Modeling Multivariate Financial Time Series Based on Correlation Clustering

ZHOU, Tu

A Thesis Submitted in Partial Fulfillment
of the Requirements for the Degree of
Master of Philosophy
in
Computer Science and Engineering

September 2008

Thesis/Assessment Committee

Professor LEONG Heng Wai Philip (Chair)
Professor CHAN Lai Wan (Thesis Supervisor)
Professor TAO Yufei (Committee Member)

Abstract

In financial time series analysis, identifying similar/dissimilar series is important to financial management. However how to define the similarities or dissimilarities in large amount of financial data is not an easy task. Correlation is the most widely used measure of similarity between two series for a certain period. The correlation here is a concept in broad sense, not restrict to Pearson-correlation coefficient. But it is noticed that the correlation between a pair of time series is time varying, and for different pairs the correlations follow different dynamics. To investigate the measurement of correlations and reveal the time-varying features of correlations, we exploit various methods and models for financial correlation estimation. Linear correlation coefficient and mutual information are examined as measurement of correlation between two series. The family of Dynamic Conditional Correlation (DCC) GARCH models are studies to describe the time-varying correlations between multivariate financial time series. The problem of DCC model is that all the pairs obey the same dynamics of correlations. This constraint is too tight to model all the correlations accurately.

To solve the problems of previous DCC models, we propose a novel Clustered DCC (CDCC) model where similar dynamics of correlations are clustered together. Instead of using equal dynamic for all pairs, the CDCC model only share the same dynamics within the same cluster, which highly raises the flexibility and fitness of previous DCC models. Experiments and simulation of financial applications on real world data verify the effectiveness of the whole

proposed mode over the DCC GARCH models. The CDCC model manages to describe the time-varying correlations of multivariate financial time series in one parsimonious model, while at the mean time allow different dynamics of the correlations. The diversity of correlation dynamics increases the fitness of the model. Meanwhile, utilizing clustering techniques to share the same dynamic within similar pairs avoid high number of parameters. The CDCC model is a more generalized form of dynamic correlation models. Previous DCC models can be regarded as special cases of the new model.

摘要

在金融時間序列分析中，識別相同的/相異的序列對于金融管理很重要。但是要在大量金融數據中找出並且評測相似度/相異度並不容易。相關度最常用于測量一段時間內兩個序列的相似性。這裡的相關度是廣義的，並不局限于通常所說的Pearson相關系數。然而，相關度本身也會隨時間發生變化，而且不同股票之間的相關度會有不同的變化。為考查相關度的測量方法以及隨時間變化的特性，我們檢驗了各種不同的測量金融相關度的方法和模型。線性相關系數和交互信息作為相關度的測量方法被檢驗。我們還學習了動態條件相關度多變量（Dynamic Conditional Correlation 簡稱DCC）的一系列模型來描述多變量金融時間序列之間的動態相關度。這個限制無疑太過嚴格以至于不能准確的描述所有的相關度。爲改善以前的模型，我們提出了一個新的聚合動態條件相關度模型（Clustered Dynamic Conditional Correlation 簡稱CDCC）。新模型通過聚合有類似變化的相關度，大幅度提高了之前DCC模型的靈活性和適合度。基于真實金融數據的試驗和應用證實了CDCC 模型相比于DCC和區域DCC（Block-DCC）模型的有效性。CDCC成功的以一個節儉的模型描述了多變量金融時間序列之間的相關度，並允許這些相關度有不同的動態變化。動態相關度的多樣性增加了模型的適合度。同時，利用聚合技術使類似的序列對共用同樣的動態變化避免了大量的模型參數。CDCC模型是動態相關度模型的高度概括形式。之前的DCC模型都可以看作是該新模型的特殊形式。

Acknowledgments

I would like to express my gratitude toward my supervisor Professor Chan Laiwan who guided me patiently and supported my interest in research topic of financial correlations. I am grateful to Dr. Zhang Kun for many valuable discussions and advice.

I would like to thank my colleagues and fellows who bring me a lot of laughters and joy. The time we spent together in research or life will never be forgotten. They are my lifetime friends, Ma Hao, She Yingni, Tu Shikui, Le Jiling, Xu Fang, Shi Lei.

Special thanks should be given to my parents and Chen Jianing for giving me great support and comfort.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation and Objective

In modern portfolio theory, it is no longer selecting best performed individual securities to construct a portfolio. Rather portfolios should be selected based on their overall risk-reward characteristics. Investors can reduce their exposure to individual asset risk by holding a diversified portfolio of assets. Diversification will allow for the same portfolio return with reduced risk. One of the most important indicators of diversification is the "similarity" or "dissimilarity" of securities. For example, securities in the same industry probably subject to similar economical and incidental impact. Listed companies holding by some shared stakeholder often behave similarly. However how to define the similarities/dissimilarities in large amount of financial data is not an easy task. Correlation is the most widely used measurement of similarity between two series for a certain period. The correlation here is a concept in broad sense, not restrict to Pearson correlation coefficient.

In the literature of financial correlations, there are several different correlation measurements. The Pearson correlation coefficient is the most applied as measure of dependence between financial series [3, 33, 48, 66, 74]. The Spearman's rho and Kendall's tau rank correlations are sometimes considered as alternatives [6, 25]. These three are all linear correlation measurements, which fail to capture the non-linear relationship between variables. The entropy-based mutual information, can depict not only linear correlation, but also non-linear relationship, which makes it a more general criterion to investigate relationships between variables. Mutual Information have been widely used in many aspects, such as Bioinformatics [76, 75], Database Pattern Mining [2, 40], and Feature Selection [80], etc. Recently it has addressed attention in measuring dependence between financial time series [21, 24, 37, 49]. All the above measurements have a numeric result of correlations. The copula [16, 22, 56, 67] as a model of dependence structure from marginal distribution to joint distribution gives more than just a single number. The study of copulas is a recent phenomenon in statistics. Lately it has been applied in financial modeling. There are numerous parametric families of copulas coupled to arbitrary marginal distributions [38, 63, 71] to model dependence among financial data.

These measurements of dependence can only depict or model the correlation for a certain period. However it is noticed that the correlation of a pair of time series is time varying and for different pairs the correlations follow different dynamics. The request for reliable estimates of dynamic correlation between financial variables has been the motivation for countless academic articles, conferences and finance industry. There are various models proposed to model the multivariate financial time series modeling [12, 14, 29].

Previously presented Dynamic Conditional Correlation (DCC) GARCH model families are critical milestones in modeling time-varying correlations among multivariate financial time series. They have clear computational advantages over conventional multivariate GARCH models. The drawback of original DCC model [27] is that all the correlation dynamics are constrained to be identical. Block-DCC model [8, 9] tried to release the constraint by introducing a block-diagonal structure. In Block-DCC model, the block structure, within which are similar single assets, is simply and manually determined according to the business nature. The manual grouping approach is the major problem of Block-DCC. Stocks in the same sector sometimes can perform distinctively along the time period, not to mention share the same dynamics of the correlations. Besides, grouping single stocks to share same dynamics of correlations is not reasonable in that correlation is a pair-wise concept. Copula-based DCC model [25, 61, 62, 68] extends the original DCC model to arbitrary multivariate distribution, such as Student copula and Clayton copula, etc. If Gaussian marginals are assumed and Gaussian copula is used, then the Copula-DCC is identical to DCC. However the assignment of multivariate distributions is not only arbitrary but also brings in more unknown parameters. Since the statistical distribution is not our focus, we will assume multivariate Gaussian distribution when dealing with multiple time series.

The objective of our work is to investigate and model correlations and the dynamics of the correlations among multivariate financial time series.

1.2 Major Contribution

The major contribution of this thesis constitutes as follows:

- To begin with, we introduce some well recognized correlation estimation methods, namely, linear correlation (including Pearson-correlation coefficient [55], Spearman's rho and Kendall's tau), entropy-based mutual information [17] which capture also non-linear relationship, and copula function [56]. Pearson correlation coefficient is most widely used as a basic measurement in financial time series correlation estimation. We investigate and compare Pearson correlation coefficient and mutual information as a measure of correlation/dependence between two financial time series. We design simple yet effective experiments to reveal the features of different measures on both simulated data and real world financial time series. Based on the experimental result, we summarize the pros and cons of each correlation measure.

- To solve the problems of previous DCC models, we propose Clustered Dynamic Conditional Correlation Multivariate GARCH model [83] to forecast the time-varying correlations between multiple series based on past information. The equal dynamic constraint of DCC model is too tight to model real world data accurately. While the block structure of Block-DCC model only group single stocks is not reasonable for correlation, which is a concept involving a pair of variables. Our proposed model eliminates these constraints by introducing a novel cluster structure. It clusters similar dynamics of correlations based on the autocorrelations of the cross product of standardized residuals from univariate GARCH model. Even though each pair of stocks is different from others, it is convenient to consider there are groups of pairs that have close dynamics of the volatility structure. The CDCC model highly raises the flexibility and fitness of previous DCC models. It allows different dynamics within one model, at the meantime, it shares the same dynamics within the same cluster, which maintain the parsimoniousness of DCC model. The previous DCC model families including DCC, Generalized DCC [26] and Block-DCC [8] can all be regarded as special cases of CDCC model.

- To verify the proposed model, we compare the CDCC model with DCC and Block-DCC model in terms of out of sample Quasi Maximum Likelihood and Box-Pierce Q statistic test result. The computational cost of CDCC model with different cluster numbers is analyzed. We also conduct financial applications utilizing the forecasting correlations of CDCC model. Portfolio selection application is carried out on a portfolio composed by diverse assets in Hong Kong stock market. The useful risk management measure Value at Risk is estimated based on 10 different world-wide indexes. Our model achieves considerable improvement over previous models in both applications, which demonstrate the effectiveness of CDCC model in real world financial applications.

1.3 Thesis Organization

This thesis is organized as follows: Chapter 2 introduce and investigate several correlation/dependence measurements to estimate the relationships between financial time series. In chapter 3, after a brief review of multivariate DCC GARCH model and extended Block-DCC GARCH model, the novel Clustered DCC model is proposed. We adopt Maximum Likelihood Estimation to estimate model parameters and Box-Pierce Test is introduced for model evaluation. Experimental results and financial applications on real world stock data in chapter 4 are followed by. Finally, a conclusion will be given in chapter 5.

Chapter 2

Measurement of Relationship between financial time series

The need to detect and properly measure relationship and dependence between financial time series is an essential task in economic applications. The most commonly used measurements are convenient functions of correlation motivated by linear relations. These measures tend to fail when they face non-linear, non-Gaussian processes. Shannon's mutual information function has been increasingly utilized in the literature, see [21, 37, 40]. The goal of this section is to find and investigate proper measures for relationship of financial time series. Widely used linear correlation coefficient and mutual information are discussed. The newly employed Copula function is also introduced. We design several experiments to examine the advantage and disadvantages of different measurements.

2.1 Linear Correlation

Linear correlation indicates the strength and direction of a linear relationship between two random variables. In the following we introduce several popular correlation coefficients.

2.1.1 Pearson Correlation Coefficient

The best known is the Pearson product-moment correlation coefficient, obtained by dividing the covariance of two variables by the product of their standard deviations. It can be expressed in mathematical formula as

$$r_{x,y} = \frac{cov(x,y)}{\sigma_x \sigma_y} = \frac{E((x-\mu_x)(y-\mu_y))}{\sigma_x \sigma_y} \tag{2.1}$$

Between two random variables X and Y with expected values μ_x, μ_y and standard deviations σ_x, σ_y.

If there is a sample of X and Y sized n, written as x_i, y_i where $i = 1, 2, \ldots, n$. The Pearson coefficient used to estimate the correlation of the two series is defined as

$$r_{x,y} = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n\sum x_i^2 - (\sum x_i)^2}\sqrt{n\sum y_i^2 - (\sum y_i)^2}} \tag{2.2}$$

The value of $r_{x,y}$ is in range $[-1, 1]$. The signs are used for positive linear correlations and negative linear correlations. If series X and Y have a strong positive linear correlation, $r_{x,y}$ is close to $+1$. This means that when x changes, y changes in the same direction. If the series have a strong negative linear correlation, $r_{x,y}$ is close to -1. When there is no linear correlation or a weak linear correlation, the coefficient is close to 0.

The Pearson correlation coefficient can be computed incrementally [69], which means it does not have to save all the memories of previous observations.

2.1.2 Rank Correlation

Rank correlation [42] is the study of relationships between different rankings on the same set of items. It deals with measuring correspondence between two rankings, and assessing the significance of this correspondence. Besides Pearson correlation coefficient, Spearman's rho and Kendall's tau are also applied in previous financial correlation literature [6].

Spearman's rank correlation coefficient or Spearman's ρ is a non-parametric measure of correlation. It assesses how well an arbitrary monotonic function could describe the relationship between two variables without making any assumptions about the frequency distribution of the variables. Spearman's ρ is simply a special case of Pearson coefficient in which two sets of data are converted to rankings before calculating the coefficient. If there are no tied ranks, then ρ is given by:

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2-1)}$$

where d_i is the difference between the ranks of corresponding values x_i and y_i, and n is the number of values in each data set.

Kendall's tau rank correlation coefficient or simply Kendall's τ is also a non-parametric statistic used to measure the degree of correspondence between two rankings and assessing the significance of this correspondence. Kendall tau coefficient is defined as:

$$\tau = \frac{2P}{\frac{1}{2}n(n-1)} - 1 = \frac{4P}{n(n-1)} - 1 \tag{2.3}$$

where n is the number of items, and P is the sum of the number of items ranked after the given item by both rankings.

2.2 Mutual Information

Mutual information of two random variables is a quantity that measures the mutual dependence of the two variables. In this section, we give a brief review of information theory, Shannon entropy and mutual information. This information theoretic concept has recently addressed attention in measuring nonlinear dependence between financial time series [21, 24, 37, 49].

Information Entropy

Information entropy is a measure of uncertainty of a random variable. The entropy of a continuous distribution, with probability density function (pdf) p_X of the random variable X is defined by [17, 72]:

$$H(X) = -\int p_X log\ p_X(x)dx. \tag{2.4}$$

If we have two variables X and Y, where the pdf of Y is p_Y and $p_{X,Y}$ is the joint pdf, the joint entropy is given by:

$$H(X,Y) = -\int\int p_{X,Y}(x,y) log\ p_{X,Y}(x,y)dxdy. \tag{2.5}$$

The conditional entropy is defined by:

$$\begin{aligned} H(Y|X) &= H(X,Y) - H(X) \\ &= -\int\int p_{X,Y}(x,y) log \frac{p_{X,Y}(x,y)}{p_X(x)} dxdy, \end{aligned} \tag{2.6}$$

or, in a similar way:

$$H(X|Y) = H(X,Y) - H(Y).$$

Mutual Information

The mutual information denoted as $I(X;Y)$ is defined by [17]

$$\begin{aligned} I(X;Y) &= H(Y) - H(Y|X) \\ &= H(X) - H(X|Y) \\ &= H(X) + H(Y) - H(X,Y) \\ &= \int\int p_{X,Y}(x,y) log \frac{p_{X,Y}(x,y)}{p_X(x)p_Y(y)}\ dxdy. \end{aligned} \tag{2.7}$$

Since $H(Y) \geq H(Y|X)$, we have $I(X;Y) \geq 0$, with the equality if and only if the two variables X and Y are statistically independent, i.e. $p_{X,Y}(x,y) = p_X(x)p_Y(y)$. So the mutual information gives a measure of dependence of two random variables or distribution.

Note the mutual information $I(X;Y)$ is never larger than any of the individual entropies.

$$I(X;Y) \leq min\{H(X), H(Y)\}$$

There is a consensus in literature [18, 24, 36, 49] that a good measure of dependence for a pair of random variables **x** and **y** should be required to satisfy the following six "ideal" properties.

1. It is well defined for both continuous and discrete variables
2. It is normalized to zero if **x** and **y** are independent, and lies between -1 and $+1$.
3. The modulus of the measure is equal to 1 if there is a measurable exact nonlinear relationship between the random variables.
4. It is equal to or has a simple relationship with the (linear) correlation coefficient in the case of a bivariate normal distribution.
5. It is metric, that is, it is a true measure of "distance" and not just divergence.
6. The measure is invariant under continuous and strictly increasing transformations.

Global Correlation Coefficient

The mutual information defined in equation (2.7) satisfies some properties of a good measure of dependence described above. In order to satisfy the properties 2. and 4. it is desirable to define a measure that can be compared to the linear correlation coefficient. In (2.7), we have $0 \leq\ I(X;Y)\ \leq\ +\infty$, which is difficult comparing between different samples. In this regard, [18, 37] transforms the mutual information into a standard measure, global correlation coefficient, as:

$$\lambda = \sqrt{1 - e^{-2I(X;Y)}}. \tag{2.8}$$

This function $\lambda(X;Y)$ captures the overall dependence, both linear and nonlinear between X and Y. The measure varies between 0 and 1 being directly comparable to the linear correlation coefficient. $\lambda = 0$ if and only if $I(X;Y) = 0$, that is, X and Y are independent, and $\lambda = 1$ if they are functionally related.

It is shown in [65] that, if the empirical joint distribution of X, Y is normal distribution, then the mutual information can be calculated by:

$$I(X;Y) = \frac{1}{2}\, log(1 - r^2) \qquad (2.9)$$

where r is the linear correlation coefficient between X, Y, so that $\lambda = |r|$. Because normal distribution is a "linear" distribution, in the sense that the linear correlation coefficient captures the overall dependence.

2.2.1 Approaches of Mutual Information Estimation

The difficulty to computing mutual information from sample data is that the underlying probability density function is unknown. The estimation method is extensively studied in the literature [20, 43]. There are three different methods to estimate mutual information: [19, 24]

- Histogram-based estimator;
- Kernel-based estimator;
- Parametric methods

The kernel-based estimators have too many adjustable parameters such as the optimal kernel width and the optimal kernel form, and a non-optimal choice of those parameters may cause a large bias in the results [36]. Furthermore, this kind of estimators can only deal with bivariate distributions. For the application of parametric methods one needs to know the specific form of the stochastic process.

The histogram-based estimators are the most straightforward and widely used. Due to the simplicity of computation and no need to adjust a lot of parameters, we will only focus on the histogram-based estimation in this paper.

In essence, histogram-based approach is to do partitioning in order to discretize continuous values [52, 53]. They can be divided into two groups:

- A. Equidistant cells, is to partition the data into equal distant intervals;
- B. Equiprobable cells, means partition the data in such a way that each interval will retrieve approximately the same number of data points.

The methods are be formulated in the following rules.

Method A:

1. Let R^d be the initial one-cell partition;
2. A subpartition of all cells into a^{td} subcells can be obtained by dividing each edge into a equidistant intervals;
3. Stop the subpartitioning of a cell if the vectors of random variables $\boldsymbol{x}$ and $\boldsymbol{y}$ are uniformly distributed.

Method B:

1. Let R^d be the initial one-cell partition;
2. A subpartition of all cells into a^{td} subcell can be obtained by dividing each edge into a equiprobable intervals;
3. Stop the subpartitioning of a cell if the vector of random variables $\boldsymbol{x}$ and $\boldsymbol{y}$ are conditionally independent on it.

The number a of equiprobable intervals is arbitrary. However a large a will complicate unnecessarily the calculus. In order to simplify computation, we choose $a = 2$ in computation.

The equiprobable method presents some advantages in the flexibility and adequacy for data. It sustains invariance of mutual information under one-to-one transformation of its component variables. [19]

$$I((f_1(X_1), \ldots, f_{da}(X_{da}), (f_{da+1}(X_{da+1}), \ldots, f_d(X_d))) = I((X_1, \ldots, X_{da}), (X_{da+1}, \ldots, X_d)) \tag{2.10}$$

where f_i denotes bijective transformations. This invariance is supported by experimental data in section 2.4.

It is pointed out in [54] that the histogram-based statistic to estimate mutual information will suffer from:

1. variance
2. bias caused by the finite number of observations
3. bias caused by quantization
4. bias caused by the finite histogram

The consideration of these factors depends on the empirical application, the number of observations, the configuration of the histogram cells and the smoothness of the probability density function.

The finite size of observation sample cause systematic errors of the estimation of mutual information. According to [76]

$$\Delta I(X;Y) \leq I_{observed} - I(X;Y)_{true} = \frac{M_{xy} - M_x - M_y + 1}{2N}$$

where M_x, M_y and M_{xy} denote the number of discrete states with nonzero probability. N is the number of total samples. With $M = M_x = M_y$ we have

$$I_{observed} \approx I + \frac{(M-1)^2}{2N}.$$

2.3 Copula

A copula is used as a general way of formulating a multivariate distribution in such a way that various general types of dependence can be represented. As mentioned by Nelsen in [56], the study of copulas is a recent phenomenon in statistics. Hence the adoption of copulas in empirical finance is even new. The earliest paper to propose the use of copulas theory in the analysis of economic problems was [25] in 1999. In order to understand the copulas, consider two random variables X and Y with marginal distributions $F(x) = Pr(X < x)$ and $G(y) = Pr(Y < y)$ and joint distribution function $H(x, y) = Pr(X < x, Y < y)$. All the distribution functions belong to the interval $[0, 1]$.

A two-dimensional copula is a function $C : [0, 1]^2 \to [0, 1]$, having three properties:

1. $C(u, v)$ is increasing in u and v.
2. $C(0, v) = C(u, 0) = 0, C(1, v) = v, C(u, 1) = u$
3. $\forall u_1, u_2, v_1, v_2$ in [0,1] such that $u_1 < u_2$ and $v_1 < v_2$ we have $C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0$.

If we set $u = F(x)$ and $v = G(y)$, then $C(F(x), G(y))$ yields a description of the joint distribution of x and y. If u and v are independent, then $C(u, v) = uv$.

Sklar's theorem:

Let H be a joint distribution function with margins F and G. Then, there exists a copula C such that for all real numbers x, y

$$H(x, y) = C(F(x), G(y)). \tag{2.11}$$

Furthermore, if F and G are continuous, then C is unique. Conversely, if F and G are distributions, then the function H defined by equation (2.11) is a joint distribution function with margins F and G. The proof is first given by Sklar in [73].

The adoption of copula as dependence measure in finance has lately addressed attention [25, 61, 68]. It is especially useful when the data set do not follow multivariate normal distribution. The usage of copula can be coupled with arbitrary distributions, the commonly used are Student t-copula, Clayton copula, and Plackett's copula etc. However the arbitrary assignment of copulas is also based on the assumption of corresponding multivariate distributions. The statistical distribution will not be our focus in this thesis.

2.4 Analysis from Experimental Data

In this section, we design several simple experiments to illustrate the pros and cons of Pearson correlation coefficient and mutual information. Both artificial examples and real world data will be used. A list of stock data are collected for experiments, downloaded from Yahoo finance[1] . The return is calculated by:

$$r(t) = \ln \frac{x(t)}{x(t-1)} \tag{2.12}$$

where $x(t)$ is the adjusted close price of day t, $r(t)$ represents the return of day t.

2.4.1 Experiment 1: Nonlinearity

Correlation coefficient is quite useful in detecting linear relationship. However, it can not capture the nonlinear relationships between two variables. Sometimes a small absolute value of linear correlation coefficient deceptively indicates weak correlation between two strongly related variables. Here we use an example of nonlinearly related artificial data to examine these measures.

The data points in Figure 2.1 has the relationship as $y = sin(x)$. It is a strong nonlinear relationship between x and y. The results of the three linear correlations are

- Pearson correlation coefficient $r = 0.0473$
- Spearman rank coefficient $\rho = 0.0238$
- Kendall rank coefficient $\tau = 0.0159$

None of the three linear correlation coefficients is statistical significant at 5% level. Apparently linear correlation statistics fail to capture the strong nonlinear relationship of variables.

[1]http://finance.yahoo.com/

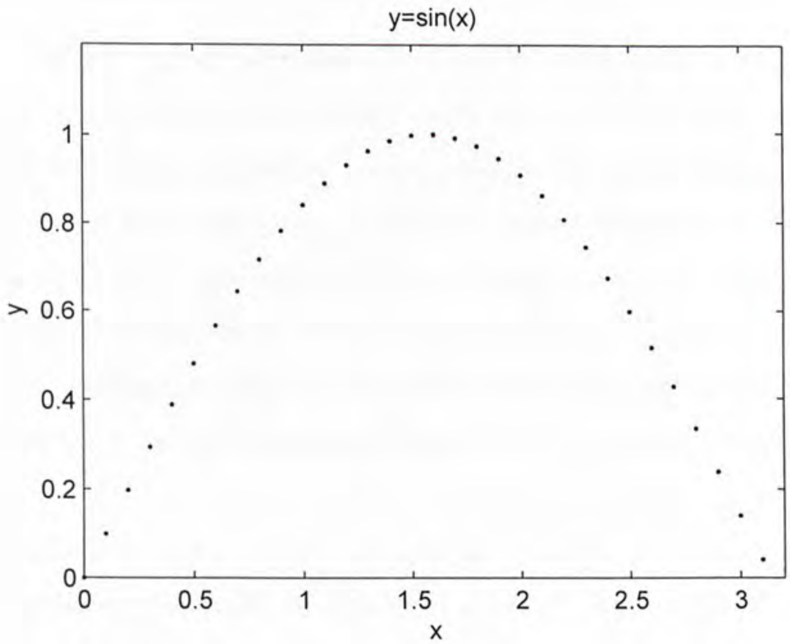

Figure 2.1: Scatter plot of two variables $x, y; y = sin(x)$

However the mutual information computed according to the methods described in last section is $I(X;Y) = 1.6523$. The details of implementation can be referred to [1]. By equation (2.8) we can transform the mutual information to global correlation coefficient $\lambda = 0.9805$. The value of λ is close to 1 correctly indicating a strong relationship between x and y. This complies with the no. 3 property of a good measurement of dependence: the modulus of the measure is equal to 1 if there is a measurable exact nonlinear relationship between the random variables.

This is an simple artificial example. In real world financial data, there is hardly any such extreme cases that are perfectly nonlinear related. But this example vividly and clearly states the weakness of linear correlation coefficients in capturing nonlinear relationships between random variables.

2.4.2 Experiment 2: Sensitivity of Outliers

In this experiment, we investigate 96 stocks of 500 trading days, which are randomly selected from Hong Kong stock market across different sections. Equiprobable method is applied to compute the mutual information, in that outlier removal will cause great diversity on mutual information computed by equidistant method. Compute the mutual information and correlation coefficient pair-wisely among the 96 stocks, sort them in ascending order respectively. We use absolute value of correlation coefficient for sorting, since ranking of original value can not demonstrate the linear relationship of the two stock data.

In total there are 4560 pairs, for each pair of stock, the mutual information and correlation coefficient has two order ranking. If the rank of the mutual information and correlation coefficient has great difference, it probably implies the two data set having some features so that these two measurements do not comply. We use a formula below to compute the difference of the ranking.

$$D = \log \frac{R_m}{R_c} \tag{2.13}$$

Here R_m indicates the rank of mutual information, R_c denotes the rank of correlation coefficient of the same pair of stock. It has some properties:

- $R_m = R_c \to D = 0$
- $R_m > R_c \to D > 0$
- $R_m < R_c \to D < 0$

Thus by finding a large absolute value of D, we can find the pairs of stocks having quite diverse ranking of mutual information and correlation. After that, we carefully examine these stocks, and find out that many of the large absolute value of D cases have outliers in the scatter plot. Thus it could be possible that the precence of outliers is one of the problems.

Stock Data Set		**Original Data**	**Outliers Removed**
0005.HK & 2878.HK	MI	0.0283	0.0218
	PCC	0.0154	0.1950
1111.HK & 1138.HK	MI	0.0128	0.0195
	PCC	0.1494	0.0510

Table 2.1: Value of mutual information and correlation coefficient

To investigate the influence of outliers, we specifically compare the mutual information and correlation coefficient of the data set of original data and after outliers removed. Since the time series data of one stock is in one dimension, we use a simple criterion to detect outliers as below.

$$if \quad |x - \mu| > 4 \times \sigma \quad set \quad x = \mu \tag{2.14}$$

Here x is the return, μ is the mean of return for 500 trading days, σ is the standard deviation. So when the absolute difference from the return value to mean is bigger than 4 times of standard deviation, the point is identified as an outlier. To maintain the consistency in the number of return, we set the outlier to the mean rather than directly delete it.

Figure 2.2 shows two examples of outlier effect. Table 2.1 is corresponding to the Figure 2.2. MI is short for mutual information while PCC stands for Pearson correlation coefficient. The original mutual information and correlation coefficient of 0005.HK and 2878.HK is 0.283 and 0.0154. When we remove the outliers according to the criteria declared above, the correlation is increases dramatically. But the mutual information only decreases slightly. In other words, the correlation coefficient was computed biased in the first place due to the presence of outliers. For the second case, after removing the outliers, the Pearson correlation coefficient drops severely, and mutual information also increases a little. Since the mutual information and Pearson correlation coefficient may have different scales, we give the histograms for reference in Figure 2.3. The sample of mutual information values has mean equals to 0.0257 and standard deviation is 0.0391. The sample of Pearson correlation coefficients

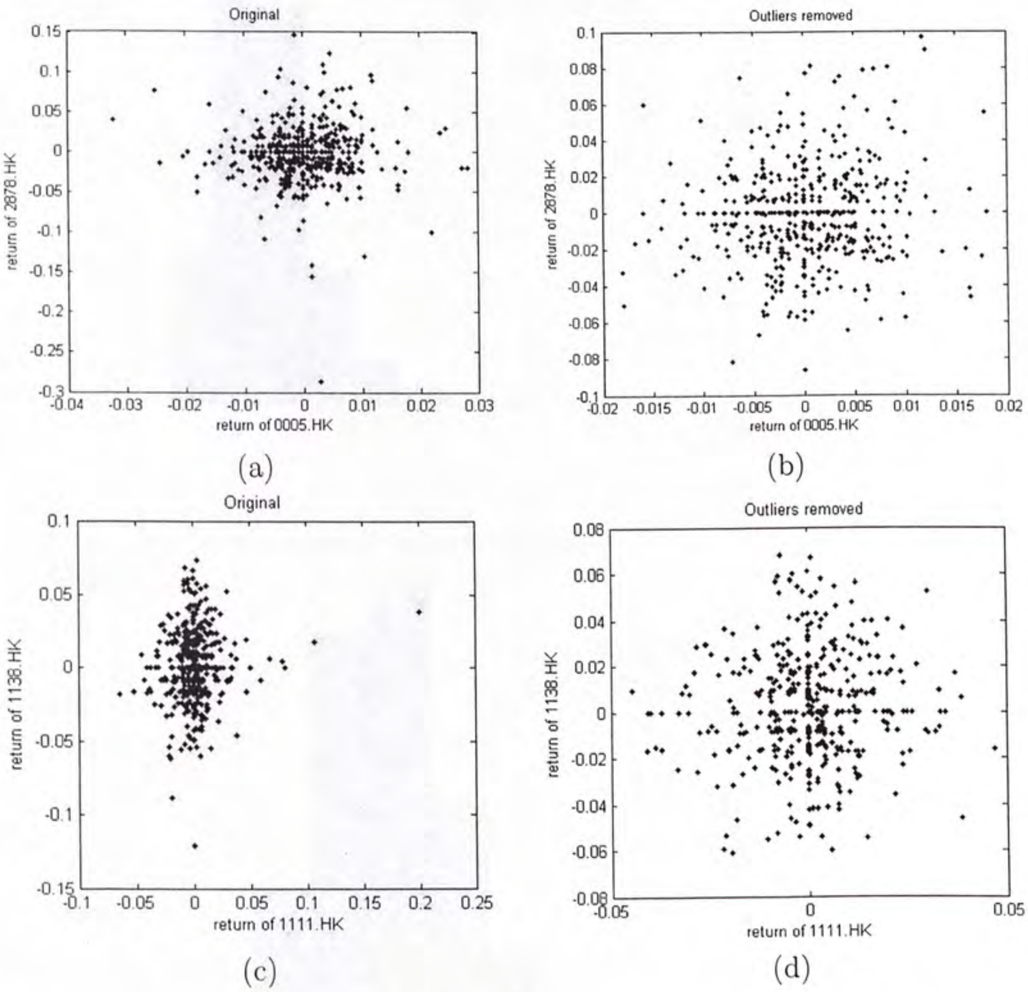

Figure 2.2: (a) (b) respectively represent the data set of 0005.HK and 2878.HK with outliers and outliers removed; (c) (d) respectively represent the data set of 1111.HK and 1138.HK with outliers and without outliers

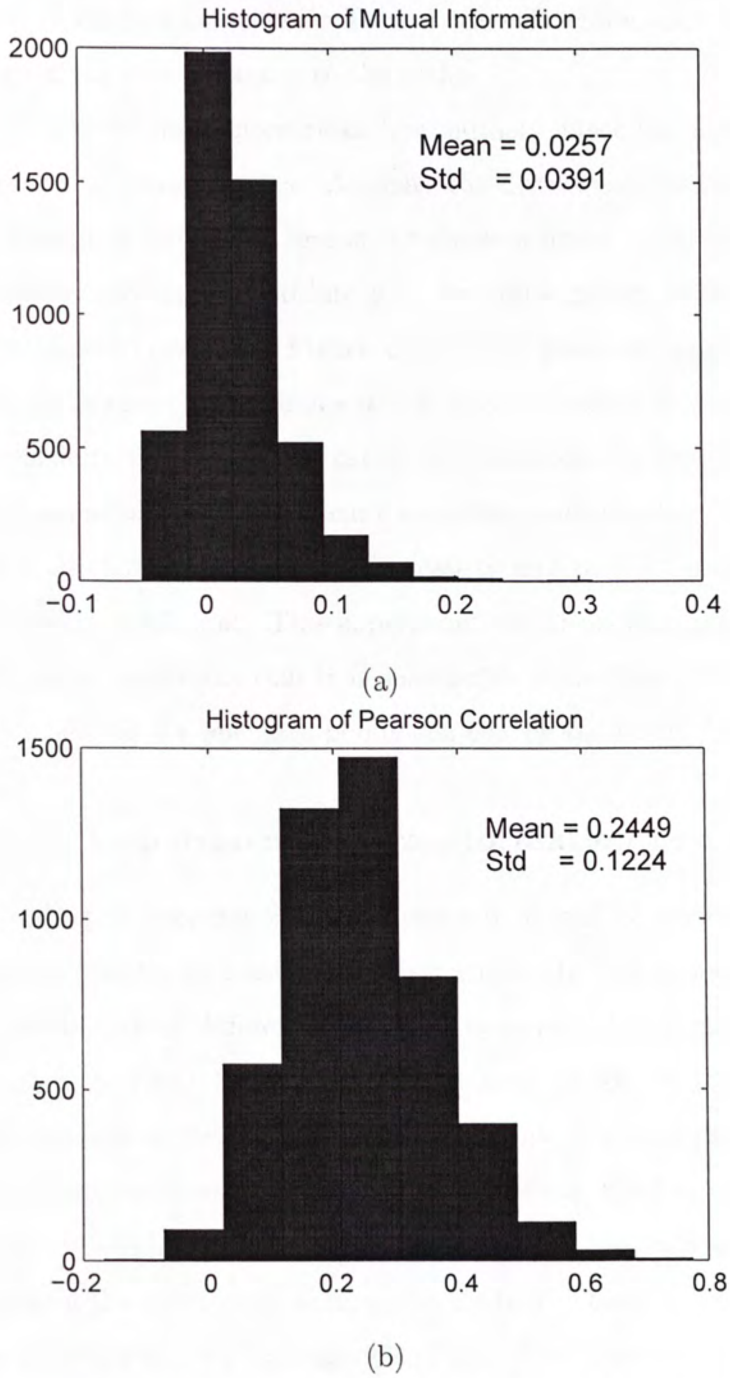

Figure 2.3: (a) Histogram of mutual information; (b) Histogram of Pearson correlation coefficients

has mean 0.2449 and standard deviation 0.1224. The change of Pearson correlation of the two examples in Table 2.1 is still significantly larger than mutual information with reference to the scales.

This example demonstrates how outliers affect the measurement of correlations of financial data. Actually the distribution of the points in Figure 2.2 has much tell us the reason. Outliers in figure (a) forms a regression line approximately parallel to line $y = -x$, thus greatly cause the entire linear correlation to decrease. Figure 2.2 (c) just gives an opposite example. The outliers line nearly in the line $y = x$, and this reflect on the increase of linear correlation. The outliers do cause estimation bias in Pearson correlation, but how the outliers influence linear correlation coefficient lies in where the outliers locate. It could cause increase, decrease or no significant change to the Pearson correlation coefficient. This experiment reveal another problem of the linear correlation coefficient that it is susceptible to outliers. As a measurement, it is not reliable if a few data points can change the result dramatically.

2.4.3 Experiment 3: Transformation Invariance

According to property 6, a good measure should be invariant under continuous and strictly increasing transformations. In this experiment, we examine the invariance of different correlation measures. 100 stocks are randomly selected from Hong Kong stock market each of 300 trading days. Therefore there should be 4950 samples of correlations. For each pair, we compute the mutual information using equiprobable method, the Pearson correlation coefficient r, Kendall's tau, and Spearman's rho. Then we transform the original return of the stocks into normal distribution. This is usually done by ranking the elements strictly increasing and applying in inverse cumulative distribution function of normal distribution. The transformation is continuous and strictly increasing. All the correlation measures are computed based on the

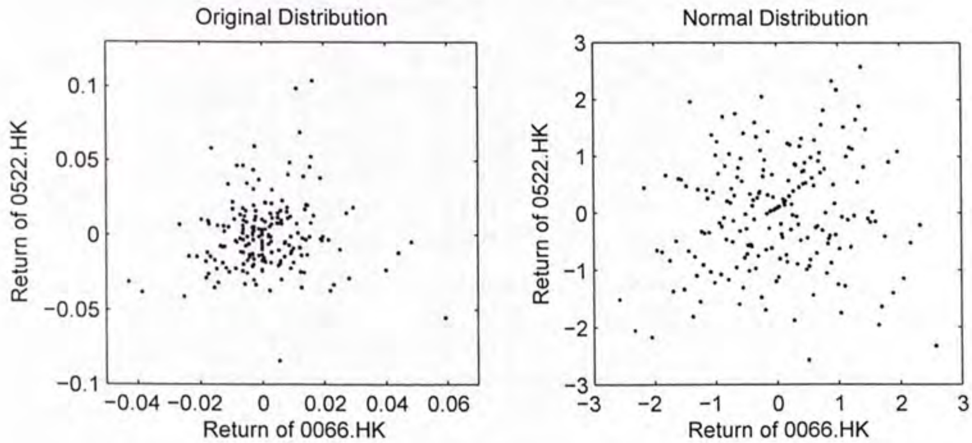

Figure 2.4: Original distribution and transformed normal distribution of 0066.HK & 0522.HK

transformed normal distributed data set again.

Figure 2.4 gives an example of the original distribution and transformed normal distribution of the returns from 0066.HK and 05222.HK. The left figure is the original distribution of return observations, the right one is the transformed normal distribution.

The result of the correlation measures of original and normal distribution is shown in Table 2.2. The symbols in the first row respectively stand for mutual information, Pearson correlation r, Kendall's τ, and Spearman's ρ. "Original" represent the original distributed samples. "Norma" indicates transformed normal distributed samples. "Absolute Diff" represents the value of absolute difference between original and normal samples. "Mean" indicates the mean value and "std" stands for standard deviation of the samples. The "SSD" means the sum of squared difference of all the samples. The "Percentage" of "mean" is computed by the mean of absolute difference over mean value of normal samples.

The two rank correlations τ, ρ has only changed slightly under the transformation, which can be told by the "Abs Diff" statistics and "Percentage". This

		MI	Pearson r	Kendall's τ	Spearman's ρ
Original	Mean	0.1612	0.2449	0.1680	0.2385
	Std	0.0650	0.1224	0.0843	0.1168
Normal	Mean	0.1635	0.2520	0.1661	0.2373
	Std	0.0618	0.1210	0.0829	0.1166
Abs Diff	Mean	0.0188	0.0239	**0.0042**	0.0045
	Std	0.0181	0.0204	**0.0044**	0.0048
	SSD	3.2924	4.7953	**0.1767**	0.2130
Percentage	Mean	11.50%	9.48%	2.50%	**1.89%**

Table 2.2: Comparison of different measures on original distribution and transformed normal distribution

is because the rank of the values would hardly change under strictly increasing transformation, so the rank correlation coefficients will remain invariant theoretically. This experiment result confirms the transformation invariance of rank correlations. When compare the mutual information and Pearson correlation, the mean and standard deviation of absolute difference and the sum of squared difference are all smaller in mutual information. But take in account of the scale level of sample values itself, we compute the percentage of mean. The global correlation coefficient is slightly better than mutual inforamtion. However according to formula (2.10), mutual information should also remain invariant under continuous and strictly increasing transformations, but the changes shown in Table 2.2 is not consistent with this finding. The reason behind this could be the easily caused bias in the estimation of mutual information. It can be caused by finite number of observations or the variance due to transformation. The difficulty and bias in estimation of mutual information makes this measurement less attractive.

2.5 Chapter Summary

In this chapter, we briefly introduce three commonly used linear correlation coefficients for financial series, the most widely applied Pearson correlation coefficient, Kendall's tau rank correlation, and Spearman's rho rank correlation. Due to the lacking in capturing nonlinear relationship of linear correlations, we consider the newly adopted mutual information in measuring dependence of financial variables.

After a breif review on the information theoretic concept, including entropy, mutual information and transformed mutual information, we also introduce approaches of estimating mutual information. We focus on histogram-based estimator and display its sub groups and properties. In addition, the copula function is introduced to complete the literature in financial dependence measurement.

To investigate the properties of different correlation measures, we design several experiments based on both artificial data and real world financial time series. By synthetic nonlinearly related data, we demonstrate the weakness of linear correlation coefficient in measuring nonlinear relationship. A mass investigation of mutual information and correlation coefficient reveals the influence of outliers to the measurement results. Different locations of outliers will cause significant increase or decrease to Pearson correlation. Thus Pearson correlation is not stable when outliers are present. An experiment of transformation invariance based on a large amount of stocks demonstrates that rank correlations are invariant under continuous and strictly increasing transformation. The Pearson correlation has larger changes than rank correlations under transformation. But surprisingly, the mutual information has changed even larger than Pearson correlation with reference to their own scales. The difficulty and easily caused bias in estimating mutual information make this measure less attractive.

There are advantages and disadvantages about Pearson correlation coefficient as well as mutual information. The biggest advantage of Pearson correlation coefficient in financial data set is that there are already a set of established theories and applications based on the bivariate normal assumption, such as Markowitz portfolio theory, variance-covariance method of Value at Risk, etc. It is hard to say which one is better, but only under certain circumstances. To summarize, the pros and cons of Pearson correlation and mutual information are listed as follows.

Pearson Correlation

Pros:

- Easy and efficient to compute.
- Capture positive and negative correlation.
- Can apply to established theory and applications.

Cons:

- Can not capture nonlinear dependence.
- Sensitive to outliers.

Mutual Information

Pros:

- Can capture both linear and nonlinear dependence.
- Not sensitive to outliers.

Cons:

- Difficult to compute, bias in estimation.
- Not a metric, can not display positivity or negativity of correlations.
- Intractable in large scale finance data.

Chapter 3

Clustered Dynamic Conditional Correlation Model

Correlation analysis is important to identify interacting pairs of time series across multiple time series data sets. In financial time series analysis, the correlations are always critical inputs for the common tasks of financial management. In this section, we will review the previous multivariate GARCH models especially the Dynamic Conditional Correlation (DCC) models. Based on previous DCC models we propose the new Clustered Dynamic Conditional Correlation Multivariate GARCH model.

3.1 Background Review

3.1.1 GARCH Model

First of all, we give a background introduction to famous GARCH model family. Volatility of a single financial time series is a significant risk measure. Traditionally volatility is represented by variance (or standard deviation). This measure of volatility is unconditional and does not recognize the interesting patterns in asset volatility, e.g., time varying and clustering properties. Over the last two decades various models are introduced to capture and forecast these patterns in volatility. One of the important volatility models is the autoregressive conditional heteroskedasticity (ARCH) proposed by Engle in 1982 [28], which propose a model where the volatility depends on past information. In 1986, Bollerslev [11] extends the ARCH model into generalized autoregressive conditional heteroskedasticity (GARCH), which better captures the time-varying volatility and clustering effect of a single financial time series.

Let y_t denote a real-valued discrete-time stochastic process, and ψ_t the information set of all information through time t. The GARCH(p,q) process is then given by:

$$y_t|\psi_t \sim N(0, h_t), \tag{3.1}$$

$$\begin{aligned} h_t &= \alpha_0 + \alpha_1 y_{t-1}^2 + \ldots + \alpha_q y_{t-q}^2 + \beta_1 h_{t-1} + \ldots + \beta_{t-p} h_{t-p} \\ &= \alpha_0 + \sum_{i=1}^{q} \alpha_i y_{t-i}^2 + \sum_{i=1}^{p} \beta_i h_{t-i}, \end{aligned} \tag{3.2}$$

where

$$\begin{aligned} & p \geq 0, \quad q > 0 \\ & \alpha_0 > 0, \quad \alpha_i \geq 0, \ i = 1, \ldots, q, \\ & \beta_i \geq 0, \quad i = 1, \ldots, p. \end{aligned} \tag{3.3}$$

For $p = 0$ the process reduces to the ARCH(q) process, and for $p = q = 0, y_t$ is simply white noise.

The simplest but most used GARCH model is of course the GARCH(1,1) model. The process $\{y_t\}$ is modeled by:

$$y_t = \sigma_t \varepsilon_t = \sqrt{h_t} \varepsilon_t \tag{3.4}$$

The standardized residuals ε_t is i.i.d.(0,1) and σ_t can be expressed in terms of previous y_t and σ_t.

$$h_t = \sigma_t^2 = \omega + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2 \tag{3.5}$$

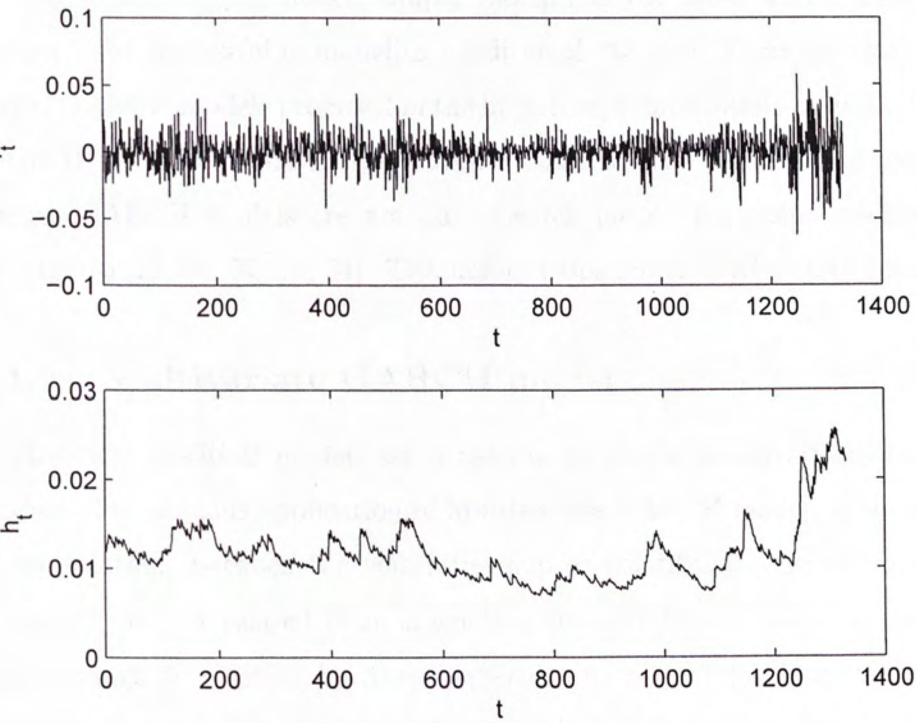

Figure 3.1: Return and Volatility of Hang Seng Index

where $\omega \geq 0, \alpha > 0, \beta > 0, \alpha + \beta < 1$ and (y_t, σ_t) is a strictly stationary solution of (3.4) and (3.5). h_t is so called the volatility of the time series.

When modeling stock return series, y_t is always the mean-corrected return process. To give a visual illustration, Figure 3.1 presents an example of the Hang Seng Index daily return series r_t and corresponding volatility h_t fitted by GARCH(1,1) model. It clearly demonstrate the time-varying and clustering effect of volatility.

The GARCH(1,1) model, simple though, is the most widely used and proven to be successful in modeling conditional variance. There are also many varied GARCH models proposed in the literature, Exponential GARCH , GJR-GARCH, Threshold GARCH, Quadratic GARCH, etc. However all kinds of diverse GARCH models are not our research focus. Interested readers are directed to [13, 35, 57, 70, 81]. Through out this paper, GARCH(1,1) is used.

3.1.2 Multivariate GARCH model

Multivariate GARCH models are a natural extension of univariate GARCH models. The obvious application of Multivariate GARCH models is the study of the relations between the volatilities and co-volatilities of several markets [7, 39, 41, 45]. A related issue is whether the correlations between asset returns change over time. Are they higher during periods of higher volatility? Are they increasing in the long run, probably because of the globalization of financial markets? Such issues can be studied directly by using a multivariate model and specifying the dynamics of covariances or correlations. Many variant models have been proposed in this respect. The general VEC model [14] simply generalize the formulation of $\boldsymbol{H}_t$. The BEKK model [29] is a special case of VEC model. The difficulty when estimating a VEC or BEKK model is the high number of unknown parameters, even after imposing several restrictions. It is thus not surprising that these models are rarely used when the number of series is larger than 3 or 4.

In 1990, Bollerslev proposed Constant Conditional Correlation model (CCC) [12] in which the conditional corrleations are constant and thus the conditional covariances are proportional to the corresponding conditional standard deviations. This restriction greatly reduces the number of unknown parameters and simplifies the estimation. The CCC model is defined as:

$$\boldsymbol{H}_t = \boldsymbol{D}_t \boldsymbol{R} \boldsymbol{D}_t = (\rho_{ij}\sqrt{h_{i,t}h_{j,t}}) \tag{3.6}$$

where

$$\boldsymbol{D}_t = diag\{\sqrt{h_{i,t}}\} \tag{3.7}$$

$h_{i,t}$ can be defined as any univariate GARCH model, and

$$\boldsymbol{R} = \{\rho_{ij}\} \tag{3.8}$$

is a symmetric positive definite matrix with $\rho_{ii} = 1, \forall i$.

The original CCC model has a GARCH(1,1) specification for each conditional variance in $\boldsymbol{D}_t$:

$$h_{i,t} = \omega_i + \alpha_i \varepsilon^2_{i,t-1} + \beta_i h_{i,t-1} \quad i = 1, \ldots, N \tag{3.9}$$

This CCC model contains $N(N+5)/2$ parameters. The assumption that the conditional correlations are constant may seem unrealistic in many empirical applications. To describe the time-varying features of the correlations, Engle presented a new class of models called Dynamic Conditional Correlation (DCC) GARCH model that not only preserve the ease of estimation of the CCC model but also allows the correlations to change over time. Engle adds to the CCC a limited dynamic in the correlations, introducing a GARCH-type structure. [44, 71, 79, 82] can be referred to for detailed survey.

3.2 DCC Multivariate GARCH Models

3.2.1 DCC GARCH Model

Given N-dimension assets, the multivariate DCC GARCH(1,1) Model [27] is defined as

$$\boldsymbol{H}_t = \boldsymbol{D}_t \boldsymbol{R}_t \boldsymbol{D}_t \quad where \quad \boldsymbol{D}_t = diag\{\sqrt{h_{i,t}}\} \tag{3.10}$$

$\boldsymbol{H}_t$ is the $N \times N$ conditional covariance matrix of mean zero return series.

$$\boldsymbol{R}_t = (\boldsymbol{Q}_t^*)^{-1} \boldsymbol{Q}_t (\boldsymbol{Q}_t^*)^{-1} \tag{3.11}$$

where the conditional covariance matrix of standardized residuals $\boldsymbol{Q}_t = \{q_{t,ij}\}$ is given by:

$$\boldsymbol{Q}_t = (1 - \alpha - \beta)\bar{\boldsymbol{Q}} + \alpha \boldsymbol{\varepsilon}_{t-1} \boldsymbol{\varepsilon}_{t-1}^T + \beta \boldsymbol{Q}_{t-1} \tag{3.12}$$

Each of the assets follows a univariate GARCH process. $\boldsymbol{\varepsilon}_t$ is a vector of univariate GARCH residuals defined as $\boldsymbol{\varepsilon}_t = \boldsymbol{D}_t^{-1} y_t$. $\bar{\boldsymbol{Q}}$ is the $N \times N$ unconditional variance matrix of $\boldsymbol{\varepsilon}_t$, in line with standard univariate GARCH result.

$$\boldsymbol{Q}_t^* = \begin{bmatrix} \sqrt{q_{t,11}} & 0 & 0 & \dots & 0 \\ 0 & \sqrt{q_{t,22}} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \sqrt{q_{t,NN}} \end{bmatrix}$$

$\boldsymbol{Q}_t^*$ is a diagonal matrix composed of square root of the diagonal elements of $\boldsymbol{Q}_t$. It is introduced to ensure that $\boldsymbol{R}_t$ is correlation matrix. The typical element of $\boldsymbol{R}_t$ will be of the form

$$\rho_{t,ij} = \frac{q_{t,ij}}{q_{t,ii} q_{t,jj}}$$

Positive definitiveness of the DCC-GARCH is controlled by the correlation function and depends on parameter restrictions, namely α and β are non-negative scalar numbers satisfying $\alpha + \beta < 1$. The drawback of DCC model is that parameters α and β are scalars, so that all the conditional correlations obey the same dynamics. This constraint is too tight to model all the assets accurately.

3.2.2 Generalized DCC GARCH Model

A simple extension on DCC GARCH model was proposed by Engle [26], who suggested the following Generalized DCC trying to solve the constraints of equal dynamics for all correlations

$$\boldsymbol{Q}_t = (\boldsymbol{\iota}\boldsymbol{\iota}' - \boldsymbol{A} - \boldsymbol{B}) \circ \bar{\boldsymbol{Q}} + \boldsymbol{A} \circ \boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}_{t-1}^T + \boldsymbol{B} \circ \boldsymbol{Q}_{t-1} \tag{3.13}$$

where $\boldsymbol{\iota}$ is a vector of ones and $\circ$ is the Hadamard product of two identically sized matrices, which is computed simply via element by element multiplication. $\mathbf{A}, \mathbf{B}$ are $N \times N$ symmetric matrices each composed of $N \times (N+1)/2$ different parameters. The Generalized DCC model simply expands the parameter α, β into parameter matrices $\mathbf{A}, \mathbf{B}$, in which each pair of series has its own dynamics. The Generalized DCC model seem to solve the constraint of equal dynamics of DCC model by allowing each pair to follow different changing correlations by its own. However the full matrix generalization incurs large increase in the number of parameters, especially when N is high. It is hard to maintain tractability, which make the model unattractive.

3.2.3 Block-DCC GARCH Model

The dynamics in DCC model are constrained to be identical for all the correlations, which is regarded as unnecessary restrictions. Billio thus release the constraints by introducing a block-diagonal structure to capture the important dependence [8, 9, 10].

The idea of Block-DCC GARCH model is to group the assets based on their business nature and constrain the dynamics to be the same within each section. The model has the same formulation as (3.13). But the parameter matrices have certain patterns. N assets are manually grouped into w sets of dimension $m_1, m_2, \dots m_w$ respectively. $\boldsymbol{\iota}_j$ indicates a column vector of ones of j dimension. Therefore the parameter matrix $\mathbf{A}$ is formed as (3.14). Similarly can matrix $\mathbf{B}$ be expressed.

$$\boldsymbol{A} = \begin{bmatrix} \alpha_{11}\boldsymbol{\iota}(m_1)\boldsymbol{\iota}(m_1)' & \alpha_{12}\boldsymbol{\iota}(m_1)\boldsymbol{\iota}(m_2)' & \dots & \alpha_{w1}\boldsymbol{\iota}(m_1)\boldsymbol{\iota}(m_w)' \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{w1}\boldsymbol{\iota}(m_w)\boldsymbol{\iota}(m_1)' & \alpha_{w2}\boldsymbol{\iota}(m_w)\boldsymbol{\iota}(m_2)' & \dots & \alpha_{ww}\boldsymbol{\iota}(m_w)\boldsymbol{\iota}(m_w)' \end{bmatrix} \tag{3.14}$$

To illustrate the format clearly, a specific example is given in (3.15) with $N = 3, w = 2$ and $m_1 = 2, m_2 = 1$. For a DCC GARCH(1,1) model, only $w + w(w-1)/2$ parameters is needed for each parameter matrix.

$$\left[\begin{array}{cc|c} \alpha_{11} & \alpha_{11} & \alpha_{12} \\ \alpha_{11} & \alpha_{11} & \alpha_{12} \\ \hline \alpha_{12} & \alpha_{12} & \alpha_{22} \end{array}\right] \tag{3.15}$$

The problem of Block-DCC model lies in the manual sectorial allocation approach. For example, in [9] the Italian Mibtel general index is grouped into three major sectors: Industrials, Services and Finance. This kind of grouping request a priori knowledge of the stock market and the criteria of sections are subjective. Sometimes the sections are simply based on the business nature of the companies. Therefore it would be difficult to categorize for those listed companies which cover various areas including finance, utility and industry. However even if the categorizing is easy, the grouping method is not reasonable across all assets. Stocks in the same sector can perform distinctively along the time period. In addition, correlation is a concept involving a pair of stocks. It would be unreasonable to group similar single stocks to share the same dynamics.

3.3 Clustered DCC GARCH Model

To solve the problems of previous DCC models, we propose a novel Clustered DCC model which cluster similar correlations between stock pairs together. The correlations are observed to change similarly due to certain impact [61]. For example, the correlations between US Dollar and other currencies drop when USD appreciates, raise when USD depreciates. Even though the correlation of each pair of stocks is different from others, it is convenient to consider there are groups of pairs that have close dynamics of the correlations. By clustering the stock pairs, the correlation model can share the same dynamics within the same cluster. Thus close correlations between stock pairs are sharing the same parameters. This extension allows flexibility of the DCC model without introducing large amount of parameters.

To clarify our model, a framework is depicted in Fig. 3.2. The rectangles represent data while rounded rectangles indicate operations. In phase 1, each single asset is modeled by univariate GARCH, and the standardized residuals $\boldsymbol{\varepsilon}_t$ of all the series are obtained. In phase 2, Minimum Distance Estimation (MDE) is applied on $\boldsymbol{\varepsilon}_t$ to obtain the parameters matrix $\boldsymbol{A}_0$ and $\boldsymbol{B}_0$, which are then clustered to form a cluster structure ζ. These are detailed discussed in this section. $\boldsymbol{\varepsilon}_t$ and ζ are then plugged in Phase 3 to estimate parameter matrix $\mathbf{A}$ and $\mathbf{B}$. Both univariate GARCH model and multivariate correlation model are employing Maximum Likelihood Estimation, which will be given in section 3.4.1.

The major difference of our model from DCC GARCH models is the extra part Phase 2. In original DCC model, the standardized residuals from each univariate GARCH model are directly used to estimate parameters of DCC model. In the proposed Clustered DCC (CDCC) model, there is an extra step 2, in which the cluster structure is employed to construct clustered parameter matrices. The parameters are then estimated applying Quasi Maximum

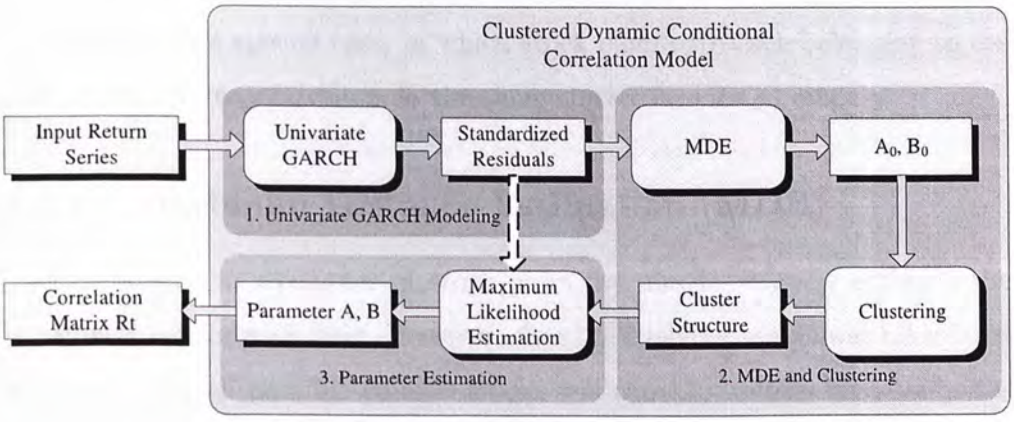

Figure 3.2: CDCC Model Framework. Phase 1 Univariate GARCH modeling; P hase 2 produce a cluster structure based on the standardized residuals obtained from Phase 1; Phase 3 estimate the parameters of the model and output the correlation matrix R_t

Likelihood estimation with parameters clustered in certain structure.

With reference to DCC model formulation, the CDCC model is slightly different in the parameter formations. For k-cluster CDCC model, suppose the distinctive parameters used in the parameter matrix $\boldsymbol{A}$ and $\boldsymbol{B}$ are $\{\alpha_1, \dots, \alpha_k\}$ and $\{\beta_1, \dots, \beta_k\}$ respectively, the k-cluster CDCC model can be formulated as, for each stock pair (i, j)

$$R_{t,ij} = Q_{t,ij} / \sqrt{Q_{t,ii} Q_{t,jj}} \tag{3.16}$$

$$Q_{t,ij} = (1 - A_{ij} - B_{ij})\bar{Q}_{ij} + A_{ij}\varepsilon_{t-1,i}\varepsilon_{t-1,j} + B_{ij}Q_{t-1,ij} \tag{3.17}$$

where $A_{ij} = \alpha_s, B_{ij} = \beta_s$ when stock pair (i, j) is in cluster s, $s = 1, \dots, k$. The univariate process are modeled using GARCH(1,1) as in DCC model.

The CDCC model can be regarded as a highly generalization of previous DCC models. The novel clustering idea not only provides flexibility to the modeling of various multivariate series, but also unify the family of DCC models. The original DCC GARCH model is the single (the least) cluster case of our model, and the Generalized DCC GARCH model can be regarded as

$N \times (N-1)/2$ (the most) clusters case. The Block-DCC model could also be regarded as a special case, in which stock pairs with each belonging to the same category respectively is in the same cluster for CDCC model.

3.3.1 Minimum Distance Estimation (MDE)

To cluster similar dynamics of stock pairs, we aim to coarsely estimate the parameter set for each pair. However, directly applying Maximum Likelihood Estimation on all possible pairs of stocks will cause extremely high computation complexity. In contrast, MDE [4], by minimizing the Mahalanobis distance of a vector of sample autocorrelations from the corresponding population autocorrelations, provides more efficient estimation, and requires no strong distributional assumptions. For completeness we briefly review MDE here. When $\alpha + \beta < 1$, the GARCH(1,1) process can be represented as:

$$\sigma_t^2 = \sigma^2(1-\alpha-\beta) + \alpha y_{t-1}^2 + \beta\sigma_{t-1}^2 \tag{3.18}$$

The first g autocorrelations of the squared series y_t^2 are in the vector $\boldsymbol{\rho}' = [\rho_1, \rho_2, \ldots, \rho_g]$. From a realization of the process y_t, the sample autocorrelations are given by $\hat{\boldsymbol{\rho}}' = [\hat{\rho}_1, \hat{\rho}_2, \ldots, \hat{\rho}_g]$, for $t = 1, 2, ..., T$

$$\rho_k = \frac{\sum_{t=1}^{T-k}(y_t^2 - \bar{y^2})(y_{t+k}^2 - \bar{y^2})}{\sum_{t=1}^{T}(y_t^2 - \bar{y^2})^2}$$

and $\bar{y^2}$ is the sample mean.

According to [23] the autocorrelation $\boldsymbol{\rho}'$ can be derived from the parameter vector $\boldsymbol{\lambda}$ of the univariate GARCH model, here specifically α, β.

$$\rho_1 = (\alpha + \frac{\alpha^2\beta}{1-2\alpha\beta-\beta^2})$$

$$\rho_k = (\alpha + \frac{\alpha^2\beta}{1-2\alpha\beta-\beta^2})(\alpha+\beta)^{k-1} \qquad for \quad k \geq 2$$

with constrain $3\alpha^2 + 2\alpha\beta + \beta^2 < 1$.

From [5], there is convergence in distribution

$$\sqrt{T} \cdot (\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}) \Rightarrow N(0, \mathbf{C})$$

where $\mathbf{C}$ is the $g \times g$ matrix with (i,j)th element given by

$$c_{ij} = \sum_{k=1}^{\infty} (\rho_{k+i} + \rho_{k-i} - 2\rho_i \rho_k)(\rho_{k+j} + \rho_{k-j} - 2\rho_j \rho_k)$$

With the theoretical support above, the parameters of a stable GARCH(1,1) model can be estimated from the autocorrelations of the squared process.

$$\hat{\boldsymbol{\lambda}} = \arg\ min\{S\} = \arg\ min\{(\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}(\boldsymbol{\lambda}))^T \boldsymbol{W} (\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}(\boldsymbol{\lambda}))\}$$

The optimal weighting matrix is $\mathbf{W} = \mathbf{C^{-1}}$. A consistent estimator of $\mathbf{C}$ is $\hat{\mathbf{C}}$, the $g \times g$ sample counterpart of $\mathbf{C}$. with $(i, j)th$ element given by

$$\hat{c}_{ij} = \sum_{k=1}^{\infty} (\hat{\rho}_{k+i} + \hat{\rho}_{k-i} - 2\hat{\rho}_i \hat{\rho}_k)(\hat{\rho}_{k+j} + \hat{\rho}_{k-j} - 2\hat{\rho}_j \hat{\rho}_k)$$

Thus, in practice, the desired parameter vector $\boldsymbol{\lambda}$ is obtained by minimizing

$$S = (\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}(\boldsymbol{\lambda}))^T \hat{\boldsymbol{C}}^{-1} (\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}(\boldsymbol{\lambda})) \tag{3.19}$$

3.3.2 Clustered DCC (CDCC) based on MDE

Since the variance matrix $\boldsymbol{Q}_t$ in (3.17) has similar derivation as (3.18), we apply MDE on the autocorrelations of the cross product of two standardized residual series to estimate parameters α, β for each series pair (i, j).

For a pair of stock (i, j) where $i, j = 1, 2, \ldots, N \quad i \neq j$, there are totally $\mathcal{N} = N(N-1)/2$ pairs. The sample autocorrelation of lag k is defined by

$$\hat{\rho}_{ij,k} = \frac{\sum_{t=1}^{T-k} (\varepsilon_{i,t}\varepsilon_{j,t} - \overline{\varepsilon_i \varepsilon_j})(\varepsilon_{i,t+k}\varepsilon_{j,t+k} - \overline{\varepsilon_i \varepsilon_j})}{\sum_{t=1}^{T} (\varepsilon_{i,t}\varepsilon_{j,t} - \overline{\varepsilon_i \varepsilon_j})^2}$$

where $\overline{\varepsilon_i \varepsilon_j}$ is the sample mean. The α, β parameter of each pair should be obtained by applying MDE (3.19).

Cluster	**Mark**	**Entries**
1	$\diamondsuit$	(1,2) (1,3) (3,4)
2	$\heartsuit$	(1,4) (2,4)
3	$\triangle$	(2,3)

Table 3.1: Clustering Result Example

The parameter sets of total $\mathcal{N}$ pairs will form two $N \times N$ symmetric matrices $\boldsymbol{A}_0$ and $\boldsymbol{B}_0$ respectively. $\boldsymbol{A}_0$ and $\boldsymbol{B}_0$ are matrices of initial parameters without diagonal entries, which will then be used to form clusters. By clustering the $\mathcal{N} \times 2$ parameters into k clusters, we obtain a cluster structure ζ. Different from the parameter matrix of Block-DCC model in (3.14), the cluster structure does not result in neatly partitioned blocks. Entries in the same cluster usually scatter in the matrix. This is mainly due to the clustering is pair-wisely for CDCC model while the grouping of Block-DCC is based on single stocks. Here we use another specific example to illustrate the structure.

Assume we have $N = 4$ stocks, thus there will be $4 \times (4-1)/2 = 6$ pairs of stocks. Suppose the 6 pairs are clustered into 3 clusters, see Table 3.1. The three different marks represent the three clusters. Then the clusters structure denoted as ζ is:

$$\zeta = \begin{bmatrix} - & \diamondsuit & \diamondsuit & \heartsuit \\ \diamondsuit & - & \triangle & \heartsuit \\ \diamondsuit & \triangle & - & \diamondsuit \\ \heartsuit & \heartsuit & \diamondsuit & - \end{bmatrix} \tag{3.20}$$

The entries with the same mark corresponding to the entries in a same cluster will apply the same parameters of the dynamic conditional correlation model. The cluster structure is the same for both parameter matrices $\boldsymbol{A}$ and $\boldsymbol{B}$. There are no representations in the diagonal because the (i, i) pair is not in the cluster sample.

Suppose the distinctive parameters used in the k-cluster parameter matrix $\boldsymbol{A}$ and $\boldsymbol{B}$ are $\{\alpha_1, \ldots, \alpha_k\}$ and $\{\beta_1, \ldots, \beta_k\}$ respectively, the formation of CDCC model is, for each stock pair (i, j)

$$R_{t,ij} = \frac{Q_{t,ij}}{\sqrt{Q_{t,ii}Q_{t,jj}}} \tag{3.21}$$

$$Q_{t,ij} = (1 - A_{ij} - B_{ij})\bar{Q}_{ij} + A_{ij}\varepsilon_{t-1,i}\varepsilon_{t-1,j} + B_{ij}Q_{t-1,ij} \tag{3.22}$$

where $A_{ij} = \alpha_s, B_{ij} = \beta_s$ when stock pair (i, j) is in cluster s, $s = 1, \ldots, k$. $\bar{Q}_{ij}$ represents the unconditional covariance of stock pair (i, j).

The k distinctive parameters are filling into $N \times N$ matrix according to the structure ζ to form parameter matrices $\boldsymbol{A}$ and $\boldsymbol{B}$. The (i, i) pair are not taken into clustering, because the diagonal entry of variance-covariance matrix $\boldsymbol{Q}_t$ represents the variance of a single stock, while the off-diagonal entries indicate the covariance between two stocks, and they obey different dynamics. To maintain $\boldsymbol{Q}_t$ a covariance matrix, the diagonal of $\mathbf{A}, \mathbf{B}$ will take the same parameters in line with the subscripts. In other words, when computing $Q_{t,ij}$, parameters in A_{ii} and A_{jj} will be the same as A_{ij}. This helps maintain the positive definitiveness of $\boldsymbol{R}_t$.

The DCC GARCH model is the single (the least) cluster case of our model, and the Generalized DCC GARCH model can be regarded as $N \times (N - 1)/2$ (the most) clusters case. The Block-DCC model could also be regarded as a special case, in which stock pairs with each belonging to the same category respectively is in the same cluster for CDCC model. Our novel model highly generalizes multivariate correlation models. It is more flexible than DCC model by differentiating various dynamics among stock pairs. At the meantime, it saves considerable amount of parameters and raises efficiency compared to Generalized DCC GARCH.

3.4 Clustering Method Selection

Due to the missing of ground truth, we could not verify the clustering accuracy of real world stock series. Thus we simulated data set that is artificially clustered in certain structure, and compare the clustering result with the target to evaluate the accuracy of different clustering methods in order to choose a suitable clustering algorithm

The generation process is the reverse course of the model. First of all we initialize $N \times N$ parameters matrix $\mathbf{A}$ and $\mathbf{B}$ with randomly generated cluster structure ζ'. The unconditional covariance matrix $\bar{\mathbf{Q}}$ is obtained by randomly select $N \times N$ principle submatrix from a covariance matrix of real data. Then for each day t, $\boldsymbol{Q}_t$, $\boldsymbol{R}_t$ can be derived from (3.16),(3.17)iteratively. The standardized residuals are composed by $\boldsymbol{\varepsilon}_t = \boldsymbol{R}_t^{1/2}\boldsymbol{\epsilon}_t$, where $\boldsymbol{\epsilon}_t$ is randomly generating N dimension i.i.d.(0,1) data. By the linear transformation, $\boldsymbol{R}_t$ becomes the correlation matrix of standardized residuals $\boldsymbol{\varepsilon}_t$.

Once $\boldsymbol{\varepsilon}_t$ are generated, MDE and Clustering are applied to obtain the estimated cluster structure $\hat{\zeta}$. To evaluate the correctness of $\hat{\zeta}$ from original ζ', we define the clustering accuracy as the ratio of the correctly clustered sample number to total amount of samples. With the help of matching matrix, it is computed by the trace of matching matrix over sum of the matrix. Given an example in Table 3.2, totally 27 samples,

$$accuracy = \frac{(5+3+11)}{27} = 0.7037.$$

We regard one single process of synthetic data generation and accuracy estimation as one iteration.

Fig. 3.3 shows the accuracy of different methods with different number of stocks and clusters. Since the dimension and size of the data to be clustered is not high, only some popular yet simple clustering techniques are considered. The clustering methods compared in this experiment are: Kmeans using Euclidean distance; Kmeans using Cosine distance; Hierarchical Clustering; and

	Cluster 1	Cluster 2	Cluster 3
Cluster 1	5	3	0
Cluster 2	2	3	1
Cluster 3	0	2	11

Table 3.2: Matching Matrix Example

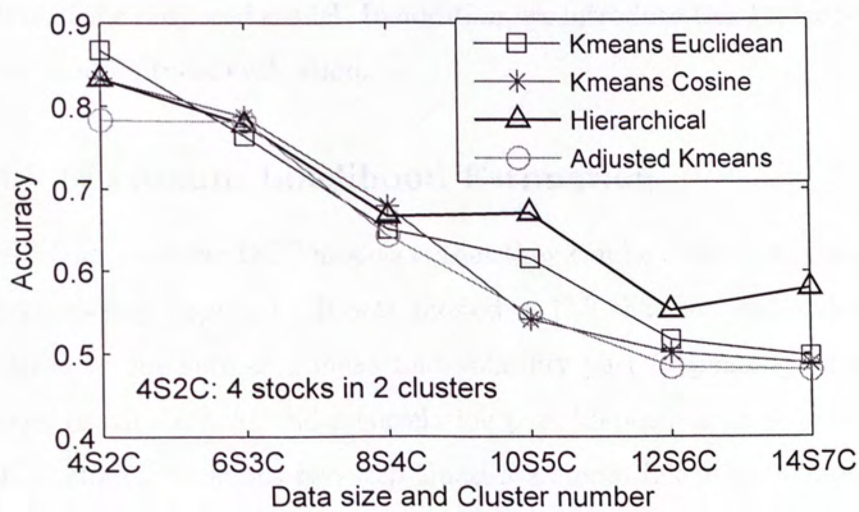

Figure 3.3: Clustering Accuracy of Different Methods and Data Sets

Adjusted Kmeans. Considering the remarkable scale difference of α and β, the Adjusted Kmeans refers to Kmeans applying on sample data that are adjusted to the same scale. For statistic purpose, we run 50 iterations for each specific stock and cluster case (corresponding to the horizontal axis tick in Fig. 3.3) to compute the average accuracy.

According to the figure, there is no significant performance difference between the algorithms when the stock and cluster numbers are small. The four lines are quite close for the first three entries. However Hierarchical clustering algorithm outperforms other algorithms when stock and cluster numbers become larger. Based on this result, the Hierarchical method is adopted in the real world data application shown in section 4.1.

3.5 Model Estimation and Testing Method

In this section, Maximum Likelihood Estimation is introduced to estimate parameters of the proposed model. In addition, we introduce Box Pierce Statistic Test for model fitness evaluation.

3.5.1 Maximum Likelihood Estimation

A useful feature of the DCC models is that they can be estimated consistently using a two-step approach. It was showed in [27] that the loglikelihood can be written as the sum of a mean and volatility part (depending on a set of unknown parameters θ_1) and a correlation part (depending on θ_2). Following the DCC model, we adopt two-step Quasi-Maximum Likelihood estimation.

Corresponding to Fig. 3.2, univariate GARCH models are estimated for each return series, in Phase 1. The standardized residuals obtained from Phase 1 and the cluster structure obtained from Phase 2 are used to estimate the parameters of the dynamic correlation. Denote the parameters of the univariate GARCH models and the parameters of the dynamic correlation by θ_1 and θ_2

respectively. Recalling that the conditional variance matrix of a CDCC model can be expressed as

$$\boldsymbol{H}_t = \boldsymbol{D}_t \boldsymbol{R}_t \boldsymbol{D}_t.$$

For N-dimensional process $\boldsymbol{y}_t$, the likelihood of the model can be written as:

$$\begin{aligned}
LogL(\theta_1, \theta_2|\boldsymbol{y}_t) &= -\frac{1}{2}\sum_{t=1}^{T}[Nlog(2\pi) + log(|\boldsymbol{H}_t|) + \boldsymbol{y}_t \boldsymbol{H}_t^{-1} \boldsymbol{y}_t^T] \\
&= -\frac{1}{2}\sum_{t=1}^{T}[Nlog(2\pi) + log(|\boldsymbol{D}_t \boldsymbol{R}_t \boldsymbol{D}_t|) + \boldsymbol{y}_t \boldsymbol{D}_t^{-1} \boldsymbol{R}_t^{-1} \boldsymbol{D}_t^{-1} \boldsymbol{y}_t^T] \\
&= -\frac{1}{2}\sum_{t=1}^{T}[Nlog(2\pi) + 2log(|\boldsymbol{D}_t|) + log(|\boldsymbol{R}_t|) + \varepsilon_t \boldsymbol{R}_t^{-1} \varepsilon_t^T]
\end{aligned}$$

In Phase 1, an inefficient but consistent esimator of parameter θ_1 can be found by replacing $\boldsymbol{R}_t$ with $\boldsymbol{I}_N$, an identity matrix of size N.

$$LogL(\theta_1|\boldsymbol{y}_t) = -\frac{1}{2}\sum_{t=1}^{T}[Nlog(2\pi) + log(\boldsymbol{I}_N) + 2log(|\boldsymbol{D}_t|) + \boldsymbol{y}_t \boldsymbol{D}_t^{-1} \boldsymbol{I}_N^{-1} \boldsymbol{D}_t^{-1} \boldsymbol{y}_t^T]$$

In Phase 3, log likelihood is conditional on the parameters estimated in Phase 1

$$LogL(\theta_2|\hat{\theta}_1, \boldsymbol{y}_t) = -\frac{1}{2}\sum_{t=1}^{T}[Nlog(2\pi) + log(\boldsymbol{R}_t) + 2log(|\boldsymbol{D}_t|) + \boldsymbol{\varepsilon}_t \boldsymbol{R}_t^{-1} \boldsymbol{\varepsilon}_t^T]$$

$\boldsymbol{\varepsilon}_t = \boldsymbol{D}_t^{-1} \boldsymbol{y}_t$ are standardized residuals obtained from univariate GARCH process. It is easier to exclude the constant terms and simply maximize:

$$LogL(\theta_2|\boldsymbol{\varepsilon}_t) = -\frac{1}{2}\sum_{t=1}^{T}[log(|\boldsymbol{R}_t|) + \boldsymbol{\varepsilon}_t^T \boldsymbol{R}_t^{-1} \boldsymbol{\varepsilon}_t]$$

Parameters obtained by Quasi-Maximum Likelihood estimation in Phase 3 are plugged into $\mathbf{A}, \mathbf{B}$ according to the cluster structure to form the parameter matrix of the model.

3.5.2 Box-Pierce Statistic Test

It is always important to evaluate the effectiveness of different kind of models. According to [31, 78, 77], compared with other alternatives, the Box-Pierce Q statistic based on the cross-product of the standardized residuals is easily computable and represents a useful diagnostic for multivariate correlation models. In this section, we introduce the Box-Pierce statistic test [64] to assess the adequacy of the proposed model.

Let $\hat{\varepsilon}_{ti}$ be the standardized residual for the i-th series, put

$$c_{t,ij} = \begin{cases} \hat{\varepsilon}_{ti}^2 - 1 & i = j \\ \hat{\varepsilon}_{ti}\hat{\varepsilon}_{tj} - \hat{R}_{t,ij} & i \neq j \end{cases}$$

where conditional correlation $\hat{R}_{t,ij} = \hat{Q}_{t,ij}/\sqrt{\hat{Q}_{t,ii}\hat{Q}_{t,jj}}$ is the estimated $R_{t,ij}$. If the multivariate conditional model fits the data, there should be no autocorrelation in $\{c_{t,ij}, t \geq 1\}$ for any fixed i, j $(i, j = 1, \ldots, N)$. Define

$$B(i, j; M) = N \sum_{k=1}^{M} \varrho_{ij,k}^2$$

$\varrho_{ij,k}$ is the sample autocorrelation of $c_{t,ij}$ at lag k. It is intuitively clear that the large value of $B(i, j; M)$ suggests model inadequacy. Thus the smaller the value of Box-Pierce test, the better the model. M is set as 5 for experiment in next section. For each stock pair (i, j), $B(i, j; 5)$ is computed.

3.6 Chapter Summary

In this chapter, we reviewed the GARCH model and multivariate GARCH model family to describe the dynamic correlations between multiple financial time series. To solve the problems of previous DCC and Block-DCC models, we propose CDCC model by clustering close dynamics of correlations together. It improves the fitness of dynamic conditional correlation estimation. MDE method is applied to obtain initial parameters which will be clustered to form a cluster structure, so that similar stock pairs share the same parameters. Due to the missing of ground truth, we generate synthetic data to observe the clustering accuracy and scalability of different clustering methods. Quasi Maximum Likelihood is employed for model estimation. In addition Box-Pierce Q statistic test is introduced to evaluate the fitness of the proposed model.

The proposed cluster structure raises the flexibility of DCC model yet still maintains the parameter parsimony of the model. Previous DCC models including DCC, Generalized DCC and Block-DCC can all be regarded as special cases of the novel CDCC model. The new model unify the DCC models in a highly generalized form. The CDCC model could also extend to copula-based CDCC model. It can be achieved by simply replacing the multivariate Gaussian distribution assumption. The Maximum Likelihood Estimation function will change accordingly. The CDCC model can be widely used in financial applications such as portfolio selection and risk management, which will be given in next chapter.

Chapter 4

Experimental Result and Applications on CDCC

In this chapter, experimental result based on DCC, Block-DCC and CDCC models are compared in terms of Quasi Maximum Likelihood and Box-Pierce statistics. We also conduct applications in portfolio selection and Value at Risk to further evaluate the CDCC model.

4.1 Model Comparison and Analysis

The data set used for simulation in this section contains the daily divident/split adjusted closing return series of 14 stocks selected from Hang Seng Index constitutes. These stocks are: { 0001.HK, 0002.HK, 0003.HK, 0004.HK, 0005.HK, 0006.HK, 0010.HK, 0011.HK, 0012.HK, 0013.HK, 0016.HK, 0019.HK, 0023.HK, 0293.HK}. The data period ranges from Jan 1990 to Sep 2007, result in 4135 observations for each stock. In order to investigate various changing dynamics in this long term period, we chop the data into five overlapping segments. For segment i,

- training data: $t = 1 + 500(i-1) \sim 1500 + 500(i-1)$
- testing data: $t = 1501 + 500(i-1) \sim 2000 + 500(i-1)$

Service		0002.HK 0003.HK 0006.HK	0013.HK 0293.HK
Finance	Real estate	0001.HK 0004.HK 0010.HK	0012.HK 0016.HK 0019.HK
	Banks	0005.HK 0011.HK	0023.HK

Table 4.1: 14 Hong Kong stocks section

	DCC	**Block**	**CDCC**				
Cluster	1	2	2	3	4	5	6
QML train	-24008.2	-24007.7	-24004.8	-24003.6	-24003	-24000.6	-23996.6
QML test	-7820.72	-7819.13	-7816.96	-7816.92	-7815.54	-7814.9	-7814.04
BP train	5.2257	5.2246	5.2224	5.2140	5.2146	5.2109	5.1993
BP test	5.0428	5.0335	4.9441	4.9437	4.9437	4.9448	4.9422

Table 4.2: Summary of different models

For Block-DCC model, the 14 stocks are manually classed into two major sectors: Service and Finance (including banks and real estate company), as shown in Table 4.1. The sections are determined based on their business nature with reference to the Block-DCC example [8].

To compare Clustered DCC (CDCC) model with original DCC and Block-DCC models, Quasi Maximum Likelihood (QML) and Box-Pierce (BP) Test are conducted on the five segments. For each segment, the model parameters are estimated using QML on training data (or in sample data), and then the estimated models are applied on testing data (or out of sample data). The average results of five segments are presented in Table 4.2 modeled by DCC, Block-DCC, and CDCC with various cluster numbers . The train and test in the first column stands for result of training data and testing data respectively.

Cluster	Time(s)	Out of sample QML	Out of sample BP
1	5.75	-7719.21	6.210
2	18.96	-7718.90	6.198
3	56.15	-7716.32	6.130
4	101.59	-7714.26	**6.104**
5	146.15	**-7712.61**	6.121
10	485.84	-7721.29	6.157
20	1430.61	-7732.36	6.218
30	2445.17	-7764.88	6.220
40	3531.26	-7758.88	6.223
50	4685.86	-7773.93	6.219
60	6010.33	-7771.69	6.180
70	7315.92	-7790.54	6.221
80	8770.12	-7801.95	6.272
90	10841.65	-7807.66	6.250

Table 4.3: Time consumption and out of sample QML, average BP value of different cluster numbers. DCC is the 1-cluster special case of CDCC

It is apparent that the CDCC model outperforms DCC and Block-DCC models in training data and testing data in terms of both QML and Box-Pierce Test. Interestingly the result of all CDCC models is nearly in direct proportion with the number of clusters. When the clusters increase, QML increases and BP value decreases, indicating higher model fitness. Therefore how to determine the number of clusters will be a compromise between model fitness and computation efficiency. However if the number of clusters exceeds certain level, the estimated parameters of some clusters would be extremely close. Then the dynamics of these clusters will not distinguish, that is why additional clusters would not be necessary.

In order to examine the computational cost with respect to the different cluster number of CDCC model, we specifically experiment and record the time used for training 500 days of the 14 stocks. Since 14 stocks totally make 91 pairs, the maximum number of clusters is 90. We apply the estimated

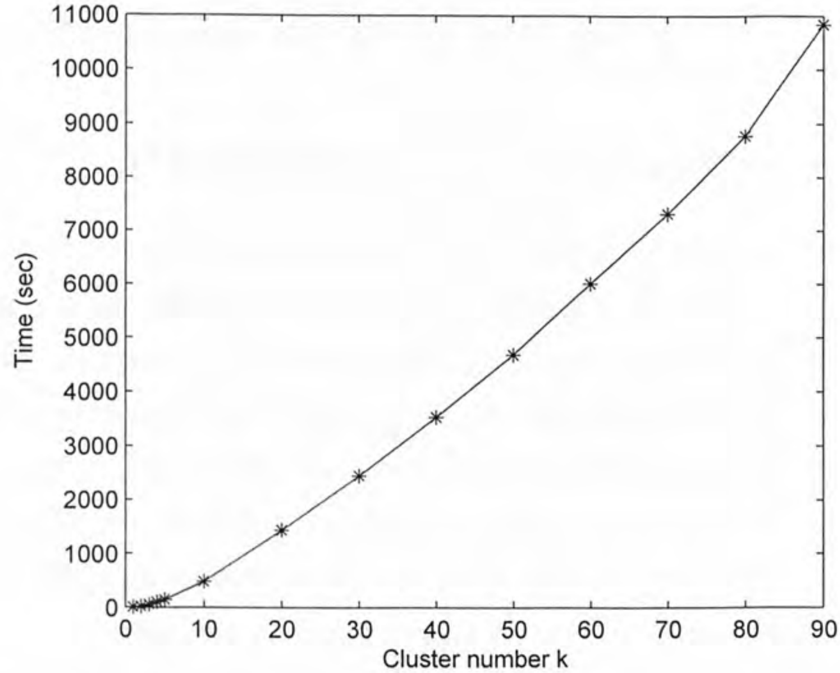

Figure 4.1: Computational Time of different cluster numbers

model on the another 500 days to compute the out of sample QML and Box-Pierce test value. The result is displayed in Table 4.3 with time consumption in seconds. All the experiments are performed on a desktop pc with 2G Hz Intel CPU and 1GB RAM.

Figure 4.1 shows the relationship of the consuming time to the number of cluster k. Based on the result of this experiment, given the same dimension of assets and same length of training data, the computational cost is a quadratic function of k. For 14 stocks of 500 days, the time used for training a 10-cluster CDCC model is nearly 8 minutes. But the time used for 90-cluster CDCC is over 3 hours, which is intolerable for the swiftly changing market. Besides, the out of sample QML and BP value begin deteriorate when the cluster number is over 20. Usually when the cluster number is too large, most of the clusters only contain one pair and all the left are in a big cluster, which could probably cause the model to over fit. The CDCC model provides an alternative of DCC model or Generalized DCC model for practitioners to determine flexibly based on the intrinsic feature of financial data and the time budget.

4.2 Portfolio Selection Application

To further verify the effectiveness of the proposed model we conduct a portfolio selection application on the same data set in section 4.1. The optimal portfolio derived from a better model should have better performance in terms of risk-return trade-off. There are various portfolio optimization approaches in literature [15, 47, 46]. However the portfolio optimization algorithm is not our focus. We just apply the classical Markowitz portfolio theory [50], which is to balance in minimizing the risk and maximizing the portfolio return.

Let $\boldsymbol{r_t}$ be the vector of asset returns with conditional mean μ_t and conditional covariance matrix $\boldsymbol{H}_t$ at time t. $\boldsymbol{\omega}_t$ is a vector of non-negative weights sum up as 1. Then the portfolio return can be computed by

$$r_t^p = \boldsymbol{\omega}_t^T \boldsymbol{r}_t$$

The risk of the portfolio is measured by the variance of the r_t^p, which can be expressed as

$$\boldsymbol{\omega}_t^T \boldsymbol{H}_t \boldsymbol{\omega}_t.$$

$\boldsymbol{H}_t$ is the variance-covariance matrix of returns obtained by equation (3.10).

The data set is the same as section 4.1. But the model training process is different. The portfolio application is conducted on the five segments of training data each size 500, based on a 1500 length moving window. For given time t, the conditional mean μ_t is estimated by moving average of historical returns, and $\boldsymbol{H}_t$ is predicted by different models using data from $t - 1500$ up to $t - 1$. The optimal portfolio weight $\hat{\omega_t}$ is obtained by solving

$$min\{\boldsymbol{\omega}_t^T \boldsymbol{H}_t \boldsymbol{\omega}_t / \boldsymbol{\omega}_t^T \mu_t\}. \tag{4.1}$$

Model	Seg 1	Seg 2	Seg 3	Seg 4	Seg 5	Mean	Annual Return
	(10^{-3})						
AvgPort	-0.860	**0.125**	-0.314	**1.040**	0.526	0.1034	2.62%
DCC	-0.286	-0.098	-0.259	0.978	0.685	0.2042	5.23%
CDCC	**-0.173**	-0.025	**-0.131**	0.989	**0.701**	**0.2722**	**7.04%**

Table 4.4: Average Daily Portfolio Return of different models

The realized portfolio return is $\hat{r}_t^p = \hat{\omega}_t{}^T r_t$. The above steps repeat and the portfolio rebalances every trading day.

Table 4.4 shows the average of realized daily portfolio return for each segment, and the mean value of 5 segments. Annual Return is computed using the mean daily return for approximately 250 trading days per year. AvgPort is portfolio with equal weight. It provides the level of average market performance as a reference. 6-cluster CDCC model is used in this experiment, and it constantly outperforms the DCC model for every segment. Thus the CDCC model achieves considerably higher annual return than DCC and equal weight portfolio.

We also notice that there are two segments the average portfolio beats our CDCC model. In segment 2, the average market has positive return but CDCC is slightly below 0. This may due to the training data. We can notice that the average return in segment 1 is a huge loss. Thus the training data and testing data have severely different features. The models estimated by previous data may not adjust to current market change. But CDCC has prevented tremendous loss in segment 1. In segment 4, the average return is very high. The models in such scenario is not effective since the overall market is up. It seems that CDCC has better performance compared to average market when the whole market is down.

From the dramatic change of average portfolio returns in the five segments, we can tell this is a long volatile period covers both bear and bull market. But the CDCC model performs more consistently than the market, it nearly triples the annual return of average market. The CDCC model also constantly beat DCC model in all the data set.

4.3 Value at Risk Application

Value at Risk (VaR) is the maximum loss not exceeded with a given probability defined as the confidence level, over a given period of time. It is widely applied in finance for quantitative risk management as an industry-wide standard [30, 34, 60]. VaR is being used for several needs; risk reporting, risk limits, regulatory capital, internal capital allocation and performance measurement.

According to [51], the VaR is a forecast of a given percentile, usually in the lower tail, of the distribution of returns on a portfolio over some period $\triangle_t$. Given some confidence level $\alpha \in (0,1)$, the VaR of the portfolio at the confidence level α is given by the smallest number l such that the probability that the loss L exceeds l is not larger than $(1-\alpha)$

$$VaR_\alpha = inf\{l \in \Re : P(L > l) \leq 1-\alpha\} = inf\{l \in \Re : F_L(l) \geq \alpha\} \tag{4.2}$$

where F_L is the distribution function of the portfolio return from time $t - \triangle_t tot$. Equivalently, we have $P(L \leq VaR_\alpha) = \alpha$ at time t. This means that we are $100(1-\alpha)\%$ confident that the loss in the period will not exceed VaR_α.

A variety methods exists for estimating VaR [32, 58, 59]. Each model has its own set of assumptions, but the most common assumption is that historical market data is our best estimator for future changes. Common models include:

1. **Historical Simulation** assuming that asset returns in the future will have the same distribution as they had in the past (historical market data),

2. **Variance-Covariance Method** assuming that risk factor returns are

always (jointly) normally distributed and that the change in portfolio value is linearly dependent on all risk factor returns,

3. **Monte Carlo Simulation** where future asset returns are more or less randomly simulated.

We apply our CDCC model on one-day period VaR application employing the Variance-Covariance method. In order to investigate versatile stock data, we collect 10 stock indexes all over global markets to form a portfolio, including AEX of Amsterdam, AORD of Australian, CAC of Paris, DAX, DJA and S&P 500 from U.S., FTSE from London, KOSPI from Korea, HSI from Hong Kong, and TWII from Taiwan. The daily returns are from Jan 2002 to Dec 2007. Take into account of the regional factors, the 10 indexes may have different market day off. To maintain the effectiveness of correlations, we remove the unaligned trading days. Thereafter all 10 data set result in 1331 aligned trading days, that is 1330 daily return observations. Since the indexes from overall the world can not be grouped to sections like the companies listed in one market. Block-DCC can not be applied in this context. Therefore we only compare DCC GARCH and CDCC model with various cluster numbers.

The first 300 training data are used to estimate the parameters of the models. Then for each t from 301 to 1330, the covariance matrix of return series $\boldsymbol{H}_t$ can be predicted by data up to $t-1$ with different models. Then the standard deviation of the portfolio is given by:

$$\sigma_p = \sqrt{\boldsymbol{\omega}^T \boldsymbol{H}_t \boldsymbol{\omega}}$$

where $\boldsymbol{\omega}$ is the vector of fixed weights of the portfolio. The normality assumption allows us to z-scale the predicted portfolio standard deviation to appropriate confidence level.

$$VaR_\alpha = z_\alpha * \sigma_p \tag{4.3}$$

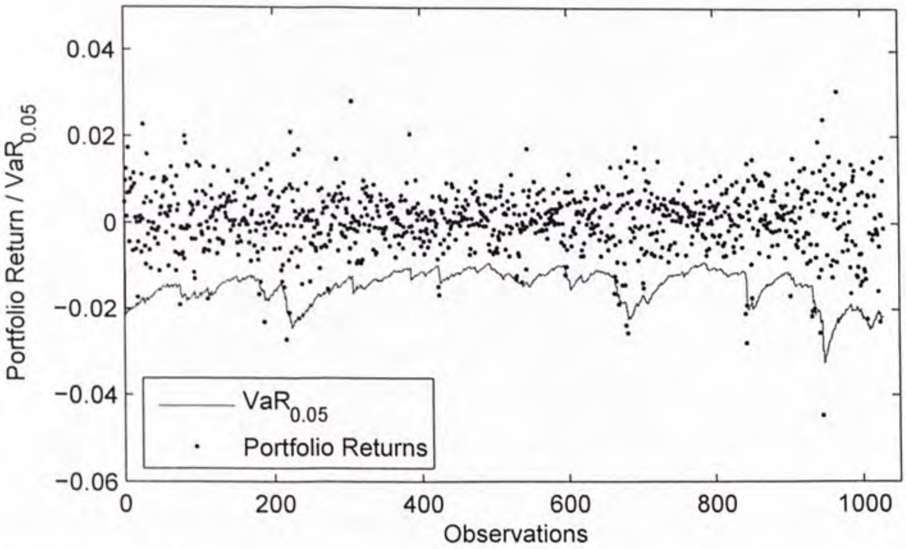

Figure 4.2: Out of sample one step ahead estimated 95% level portfolio VaR (modeled by CDCC-4) and observed portfolio returns for Portfolio 2

z_α is the critical value for corresponding confidence level. For 95% confidence level, $z = -1.645$; for 99% confidence level, $z = -2.33$.

Figure 4.2 displays an example of the estimated 95% level VaR and the portfolio returns for Portfolio 2. The data points represents the portfolio return observations, the curve is the predicted VaR loss. For any points drop below the curve represents the loss is larger than expected. Table 4.5 shows the results of VaR for three portfolios at 95% and 99% level. Portfolio 1 is equal weighted portfolio of the 10 indexes, portfolio 2 is a uneven combination, portfolio 3 is a hedge portfolio. The weight vectors are given below:

- $\omega_1 = [0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1]^T$
- $\omega_2 = [0.15, 0.15, 0.15, 0.15, 0.15, 0.05, 0.05, 0.05, 0.05, 0.05]^T$
- $\omega_3 = [0.1, 0.3, -0.2, 0.2, -0.1, 0.2, 0.1, -0.1, 0.3, 0.2]^T$

It is demonstrated that in all three different portfolios, CDCC model provides a better result in the VaR estimation for both $\alpha = 0.05$ and $\alpha = 0.01$.

Portfolio	Model	$\alpha = 0.05$	$\alpha = 0.01$
Portfolio 1	DCC GARCH	0.0437(45)	0.0165(17)
	CDCC 2-Cluster	0.0418(43)	0.0155(16)
	CDCC 3-Cluster	0.0398(41)	0.0136(14)
	CDCC 4-Cluster	**0.0389(40)**	**0.0136(14)**
Portfolio 2	DCC GARCH	0.0301(31)	0.0117(12)
	CDCC 2-Cluster	0.0292(30)	0.0107(11)
	CDCC 3-Cluster	0.0292(30)	0.0107(11)
	CDCC 4-Cluster	**0.0292(30)**	**0.0097(10)**
Portfolio 3	DCC GARCH	0.0554(57)	0.0253(26)
	CDCC 2-Cluster	0.0525(54)	0.0243(25)
	CDCC 3-Cluster	**0.0505(52)**	0.0233(24)
	CDCC 4-Cluster	0.0515(53)	**0.0224(23)**

Table 4.5: Proportion of observations (number of observations in brackets) where the portfolio loss exceeds the estimated VaR for $\alpha = 0.05$ and 0.01

4.4 Chapter Summary

A comparison between DCC, Block-DCC and CDCC models is made based on the QML and Box-Pierce test result. Theoretically the more cluster of CDCC model, the better the performance. However the computational cost also increases quadratically. The experiment result reveal that the cluster number should not exceed certain level, otherwise the model would be over fitting. The CDCC model provides an alternative of DCC model or Generalized DCC model for practitioners to determine flexibly based on the intrinsic feature of financial data and the time budget.

Furthermore we apply two financial applications utilizing the forecast correlations of the CDCC model. A simple portfolio application is conducted based on the Markowitz modern portfolio theory. The CDCC model performs more consistently than the market, it nearly triples the annual return of average market. The CDCC model also constantly beat DCC model in all the five segment data. It achieves considerably higher annual portfolio return than DCC model. A Value at Risk application is performed as a risk management metric using the variance-covariance method. This time rather than using Hong

Kong stock data, we explore international markets. The data set is 10 stock indexes from all over the world. We experiment three different portfolios, the equal weighted, unequal weighted, and hedge portfolio. The back test on the three portfolios at 95% and 99% consistently demonstrate better performance of CDCC model over DCC model.

Chapter 5

Conclusion

In this thesis, we exploit various methods and models to estimate financial correlations and the dynamics of the correlations. We briefly introduce three commonly used linear correlation coefficients for financial series, the most widely applied Pearson correlation coefficient, Kendall's tau rank correlation, and Spearman's rho rank correlation. Due to the lacking in capturing nonlinear relationship of linear correlations, we consider the newly adopted mutual information in measuring dependence of financial variables. In addition, the copula function is introduced to complete the literature in financial dependence measurement. But the experiments focus on linear correlation and mutual information, which only result in single numbers.

By synthetic nonlinearly related data, we demonstrate the weakness of linear correlation coefficient in measuring nonlinear relationship. A mass investigation of mutual information and correlation coefficient reveals the influence of outliers to the measurement results. Different locations of outliers will cause significant increase or decrease to Pearson correlation. Thus Pearson correlation is not stable when outliers are present. An experiment of transformation invariance based on a large amount of stocks demonstrates that rank correlations are invariant under continuous and strictly increasing transformation. The Pearson correlation has larger changes than rank correlations under transformation. But surprisingly, the mutual information has changed even larger than Pearson correlation with reference to their own scales. The difficulty and easily caused bias in estimating mutual information make this measure less attractive.

There are advantages and disadvantages about Pearson correlation coefficient as well as mutual information. The major advantage of Pearson correlation in financial data set is that there are already a set of established theories and applications based on the bivariate normal assumption, such as Markowitz portfolio theory, variance-covariance method of Value at Risk, etc.

These measurements of dependence can only depict or model the correlation for a certain period. In the fast changing financial world, it is noticed that the correlations between a pair of assets is also time varying. And for different pairs, the correlations follow different dynamics. In order to model the time-varying features correlations between multivariate financial time series, we reviewed variant multivariate financial time series models. Previously presented Dynamic DCC GARCH model families are critical milestones in modeling time-varying correlations among multivariate time series. They have clear computational advantages over conventional multivariate GARCH models. However the constraint of equal dynamics among all stock pairs in DCC is too tight to model all the assets correctly, especially when dimension is high. The Generalized DCC model completes the diversity of different pairs, but sacrifices the efficiency in model estimation. The following proposed Block-DCC model is problematic in the manual section approach.

To solve these problems, we present a novel Clustered DCC model which extends the previous models by incorporating clustering techniques. Instead of using the same parameters for all time series, a cluster structure is produced based on the autocorrelations of standardized residuals, in which clustered entries sharing the same dynamics. MDE method is applied to obtain initial parameters which will be clustered to form a cluster structure, so that similar stock pairs share the same parameters. The proposed cluster structure raises the flexibility of DCC model yet still maintains the parameter parsimony of the model. Original DCC, Generalized DCC and Block-DCC models can all be regarded as special cases of our model. The CDCC model provides a framework of utilizing clustering in multivariate time series models. The cluster numbers in the experiment of this thesis are predefined. However the clustering method can be replaced by any state-of-art clustering techniques to automatically determine the number of clusters. The CDCC model could also extend to copula-based CDCC model. It can be achieved by simply replacing the multivariate Gaussian distribution assumption. The Maximum Likelihood Estimation function will change accordingly.

To verify the effectiveness of the whole proposed model, we performed Quasi Maximum Likelihood estimation and Box-Pierce Q statistic test to evaluate the goodness-of-fit. We also experiment on the scalability and computational cost along with the increasing of cluster numbers. According to our experiment record, the consuming time is approximately quadratic function of the number of clusters in CDCC model. For 14 stocks of 500 days, the time used for training a 10-cluster CDCC model is nearly 8 minutes. This cost is affordable even for daily estimation of the model. But the time used for 90-cluster CDCC of the same data set is over 3 hours, which is intolerable for the swiftly changing market. Besides the overly high number of clusters would probably cause the model to over fit. The CDCC model provides an alternative of DCC model or Generalized DCC model for practitioners to determine flexibly based on the intrinsic feature of financial data and the time budget. We also simulate financial applications such as portfolio selection and Value at Risk on diverse stock markets as a guidance of how CDCC model can be applied in real-world financial applications. The results of the experiments and applications demonstrate that the CDCC model outperforms the previous DCC models within reasonable time to afford.

The significant contribution of CDCC model is introducing the concept of clustering into the multivariate time series modeling. That is to cluster similar features in high dimensional data and model them in a clustered structure. The employment of clustering is not restricted to GARCH models or financial time series. The dimension of multivariate time series modeling has always been the problem. CDCC, however, as an example in financial correlation models, has provided a guidance of how to maintain computation efficiency while enhance the flexibility and fitness of multivariate time series models.

Bibliography

[1] http://www.cs.rug.nl/ rudy/matlab/doc/information.html.

[2] Periklis Andritsos, Renée J. Miller, and Panayiotis Tsaparas. Information-theoretic tools for mining database structure from large data sets. In *SIGMOD '04: Proceedings of the 2004 ACM SIGMOD international conference on Management of data*, pages 731–742, New York, NY, USA, 2004. ACM.

[3] Andrew Ang and Joseph Chen. Asymmetric correlations of equity portfolios. *Journal of Financial Economics*, 63:443–494, March 2002.

[4] Richard T. Baillie and Huimin Chung. Estimation of garch models from the autocorrelations of the squares of a process. *Journal of Time Series Analysis*, 22:631–650(20), 2001.

[5] M. S. Bartlett. On the theoretical specification and sampling properties of autocorrelated time-series. *Journal of Royal Statistical Society*, 8(1):27–41.

[6] S.M. Bartram and Y.H. Wang. Another look at the relationship between cross-market correlation and volatility. *Finance Research Letters*, 2:75–88, June.

[7] Luc Bauwens and Jeroen Rombouts. Bayesian clustering of many garch models. *Econometric Reviews Special Issue Bayesian Dynamic Econometrics*, 2003.

[8] Caporin M. Billio, M. and M. Gobbo. Block dynamic conditional correlation multivariate garch models. *Greta Working Paper*, 2003.

[9] Caporin M. Billio, M. and M. Gobbo. Flexible dynamic conditional correlation multivariate garch models for asset allocation. *Applied Financial Economics Letters*, 2:123–130, 2006.

[10] Monica Billio and Massimiliano Caporin. A Generalized Dynamic Conditional Correlation Model for Portfolio Risk Evaluation. Technical report, the Department of Economics of the Ca' Foscari University of Venice, 2006.

[11] T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31:307–327, 1986.

[12] T. Bollerslev. Modelling the coherence in short-run nominal exchange rates: a multivareiate generalized arch approach. *Review of Economic and Statistics*, 72:498–505, 1999.

[13] T. Bollerslev, R.Y. Chou, and K.F. Kroner. Arch modeling in finance: a review of the theory and empirical evidence. *Journal of Econometrics*, 52:5–59, 1992.

[14] T. Bollerslev, R.F. Engle, and J.M. Wooldridge. A capital asset pricing model with time varying covariances. *Journal of Political Economy*, 96:116–131, 1988.

[15] Laiwan Chan. Bidirectional reinforcement learning for asset allocation. In *ICONIP*, 2005.

[16] U. Cherubini, E. Luciano, and W. Vecchiato. *Copula Methods in Finance.* Wiley, New York, 2004.

[17] Thomas M. Cover and Joy A. Thomas. *Elements of information theory.* Wiley-Interscience, New York, NY, USA, 1991.

[18] Georges A. Darbellay. *Predictability: an Information-Theoretic Perspective*, chapter 18, pages 249–262. Spriger, Birkhauser, Boston, 1998.

[19] Georges A. Darbellay. An estimator of the mutual information based on a criterion for conditional independence. *Computational Statistics & Data Analysis*, 32(1):1–17, November 1999.

[20] Georges A. Darbellay and I. Vadja. Estimation of the information by an adaptive partitioning of the observation space. *IEEE Transactions on Information Theory*, 45:1315–1321, May 1999.

[21] Georges A. Darbellay and D. Wuertz. The entropy as a tool for analysing statistical dependence's in financial time series. *Physica A*, 287:429–439, December 2000.

[22] Stefano Demarta and Alexander J. McNeil. The t copula and related copulas. *International Statistical Review*, 73(1):111–129, April 2005.

[23] Zhuanxin Ding and Clive W. J. Granger. Modeling volatility persistence of speculative returns: A new approach. *Journal of Econometrics*, 73(1):185–215, July 1996.

[24] Andreia Dionisio, Rui Menezes, and Diana A. Mendes. Mutual information: a dependence measure for nonlinear time series. *Physica A*, November 2003.

[25] P. Embrechts, A. McNeil, and Daniel Straumann. Correlation: Pitfalls and alternatives. *RISK*, 12(5):69–71, May 1999.

[26] R.F. Engle. Dynamic conditional correlation - a simple class of multivariate garch. *Journal of Business and Economics Statistics*, 17:425–446, 2002.

[27] R.F. Engle and K. Sheppard. Theoretical and empirical properties of dynamic conditional correlation multivariate garch. Technical report, University of California, San Diego, CA, 2001.

[28] Robert F. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 50(4):987–1007, 1982.

[29] Robert F. Engle and Kenneth F. Kroner. Multivariate simultaneous generalized arch. *Econometric Theory*, 11(1):122–150, 1995.

[30] Robert F. Engle and Simone Manganelli. Caviar: Conditional value at risk by quantile regression. Working Paper 7341, National Bureau of Economic Research, September 1999.

[31] Wang M. Fa, J. and Q. Yao. Modeling multivariate volatilities via conditionally uncorrelated components. Technical report, Department of Statistics,London School of Economics and Political Science, 2005.

[32] Jianqing Fan and Juan Gu. Data-analytic approaches to the estimation of value-at-risk. *Computational Intelligence for Financial Engineering, 2003. Proceedings. 2003 IEEE International Conference on*, pages 271–277, March 2003.

[33] Kristin Forbes and Roberto Rigobon. No contagion, only interdependence: Measuring stock market co-movements. *Journal of Finance*, 57:2223–2261, 2002.

[34] P. Giot and S. Laurent. Value-at-risk for long and short positions. *Journal of Applied Econometrics*, 18:641–664, 2003.

[35] Lawrence R. Glosten, Ravi Jagannathan, and David E. Runkle. On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance*, 48(5):1779–1801, 1993.

[36] C. W. Granger, E. Maasoumi, and J. Racine. A dependence metric for possibly nonlinear processes. *Journal of Time Series Analysis*, 25(5):649–669, Sep 2004.

[37] Clive Granger and Jin-Lung Lin. Using the mutual information coefficient to identify lags in nonlinear models. *Journal of Time Series Analysis*, 15(4):371–384, 1994.

[38] Eric Jondeau and Michael Rockinger. The copula-garch model of conditional dependencies: An international stock market application. *Journal of International Money and Finance*, 25(5):827–853, August 2006.

[39] G.A. Karolyi. A multiavariate garch model of international transmission of stock returns and volatility; the case of the united states and canada. *Journal of Business and Economic Statistics*, 13:11–25, 1995.

[40] Yiping Ke, James Cheng, and Wilfred Ng. Mining quantitative correlated patterns using an information-theoretic approach. In *KDD '06: Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 227–236, New York, NY, USA, 2006. ACM.

[41] C. Kearney and A.J. Patton. Multivariate garch modelling of exchange rate volatility transmission in the european monetary system. *Financial Review*, 41:29–48, 2000.

[42] M. Kendall. *Rank Correlation Methods.* A Charles Griffin Book, 5th edition, Sep 1990.

[43] A. Kraskov, H. Stögbauer, and P. Grassberger. Estimating mutual information. *Physical Review*, 69(6), June 2004.

[44] Sebastien Laurent, Luc Bauwens, and Jeroen V. K. Rombouts. Multivariate garch models: a survey. *Journal of Applied Econometrics*, 21(1):79–109, 2006.

[45] O. Ledoit, P. Santa-Clara, and M. Wolf. Flexible multivariate garch modeling with an application to international stock markets. *The Review of Economics and Statistics*, 85:735–747, 2003.

[46] Jian Li and Laiwan Chan. Reward adjustment reinforcement learning for risk-averse asset allocation. *Neural Networks, 2006. IJCNN '06. International Joint Conference on*, pages 534–541, 2006.

[47] Jian Li and Laiwan Chan. Independent factor reinforcement learning for portfolio management. In *Intelligent Data Engineering and Automated Learning - IDEAL 2007*, volume 4881 of *Lecture Notes in Computer Science*, pages 1020–1031. Springer Berlin / Heidelberg, 2007.

[48] Francois Longin and Bruno Solnik. Is the correlation in international equity returns constant: 1960-1990? *Journal of International Money and Finance*, 14(1):3–26, February 1995.

[49] Esfandiar Maasoumi and Jeff Racine. Entropy and predictability of stock market returns. *Journal of Econometrics*, 107(1-2):291–312, March 2002.

[50] Harry Markowitz. Portfolio selection. *The Journal of Finance*, 7(1):77–91, March 1952.

[51] Alexander McNeil, Rüdiger Frey, and Paul Embrechts. *Quantitative Risk Management: Concepts, Techniques, and Tools.* Princeton Series in Finance. Princeton University Press, 2005.

[52] R. Moddemeijer. On estimation of entropy and mutual information of continuous distributions. *Signal Processing*, 16(3):233–246, 1989.

[53] R. Moddemeijer. An efficient algorithm for selecting optimal configurations of AR-coefficients. In A. Barbé et. al., editor, *Twentieth Symposium on Information Theory in the Benelux*, pages 189–196. Werkgemeenschap Informatie- en Communicatietheorie, Enschede (NL), May 1999.

[54] R. Moddemeijer. A statistic to estimate the variance of the histogram based mutual information estimator based on dependent pairs of observations. *Signal Processing*, 75(1):51–63, 1999.

[55] David S. Moore. *The Basic Practice of Statistics with Cdrom.* W. H. Freeman & Co., New York, NY, USA, 1999.

[56] Roger B. Nelsen. *An Introduction to Copulas.* Springer-Verlag New York, Inc., Secaucus, NJ, USA, 1994.

[57] Daniel B. Nelson. Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, 59(2):347–370, 1991.

[58] Alper Ozun and Atilla Cifter. Estimating portfolio risk with conditional joe-clayton copula: An empirical analysis with asian equity markets. *The Icfai Journal of Financial Economics*, (3):28–41, September 2007.

[59] Alper Ozun and Atilla Cifter. Portfolio value-at-risk with time-varying copula: Evidence from the americas. MPRA Paper 2711, University Library of Munich, Germany, April 2007.

[60] H.P. Palaro and L.K. Hotta. Using conditional copula to estimate value at risk. *Journal of Data Science*, 4:93–115, 2006.

[61] Andrew J. Patton. Modelling time-varying exchange rate dependence using the conditional copula. University of California at San Diego, Economics Working Paper Series 2001-09, Department of Economics, UC San Diego, June 2001.

[62] Andrew J. Patton. Estimation of copula models for time series of possibly different lengths. University of California at San Diego, Economics Working Paper Series 2002-05, Department of Economics, UC San Diego, 2002.

[63] Andrew J. Patton. Copula-based models for financial time series. OFRC Working Papers Series 2008fe21, Oxford Financial Research Centre, 2008.

[64] David A. Pierce. Distribution of residual autocorrelations in the regression model with autoregressive-moving average errors. *Journal of the Royal Statistical Society. Series B (Methodological)*, 33(1):140–146, 1971.

[65] M. S. Pinsker. *Information and Information Stability of Random Variables and Processes.* San Francisco: Holden Day, 1964.

[66] L. Ramchand and R. Susmel. Volatility and cross correlation across major stock markets.

[67] Jorn Rank. *Copulas: From theory to application in finance.* Risk Books, 2006.

[68] Michael ROCKINGER and Eric JONDEAU. Conditional dependency of financial series : an application of copulas. Working paper 723, Group HEC, February 2001.

[69] Y. Sakurai, S. Papadimitriou, and C. Faloutsos. Braid: stream mining through group lag correlations. In *SIGMOD '05: Proceedings of the 2005 ACM SIGMOD international conference on Management of data*, pages 599–610, New York, NY, USA, 2005. ACM.

[70] Enrique Sentana. Quadratic arch models. *The Review of Economic Studies*, 62(4):639–661, 1995.

[71] Brockwell A. Lehoczky J. Serban, M. and S. Srivastava. Modeling the dynamic dependence structure in multivariate financial time series. *Journal of Time Series Analysis*, 2006.

[72] C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27(3):379–423, 1948. Continued in following volume.

[73] A. Sklar. Fonctions de répartition àn dimensions et leurs marges. *Publications de Institut de Statistique de ĹUniversité de Paris*, 8:229–231, 1959.

[74] B. Solnik, C. Boucrelle, and Y.L. Fur. International market correlation and volatility. *Financial Analysts Journal*, 52:17–34, 1996.

[75] R. Steuer, C.O. Daub, J. Selbig, and J. Kurths. Measuring distances between variables by mutual information. In *27th Annual Conference of the Gesellschaft fur Klassifikation: Cottbus*, volume 12-14, 2003.

[76] R. Steuer, J. Kurths, C.O. Daub, J. Weise, and J. Selbig. The mutual information: Detecting and evaluating dependencies between variables. *Bioinformatics*, 18:231–240, 2002.

[77] Tsui A.K.C. Tse, Y.K. A multivariate garch model with time-varying correlations. *Journal of Business and Economic Statistics*, 20:351–362, 2000.

[78] Y.K. Tse and A.K.C. Tsui. A note on diagnosing multivariate conditional heteroscedasticity models. *Journal of Time Series Analysis*, 20:679–691, 1999.

[79] G.A. Vargas. An asymmetric block dynamic conditional correlation multivariate garch model. *Philippines Statistician*, 2006.

[80] G. Wang, F.H. Lochovsky, and Q. Yang. Feature selection with conditional mutual information maximin in text categorization. In *Conference on Information and Knowledge Management*, pages 342–349, 2004.

[81] Jean-Michel Zakoian. Threshold heteroskedastic models. *Journal of Economic Dynamics and Control*, 18(5):931–955, September 1994.

[82] Kun Zhang and Laiwan Chan. Efficient factor garch models and factor-dcc models. *Quatitative Finance*, 2008.

[83] Tu Zhou and Laiwan Chan. Clustered dynamic conditional correlation multivariate garch model. In *Dawak 2008: 10th International Conference on Data Warehousing and Knowledge Discovery*. Springer, 2008.